

Identification and expression patterns of 4-coumarate:coenzyme A ligase genes in *Agave*

Xu-Xia WANG¹ , Na ZHANG^{2,*} , Liang CHEN³ , Bo WANG⁴ , Jing-Yi LU⁵ , Xiao-Li HU⁴ , and Xing HUANG^{4,6,*} 

¹ School of Smart Cities and Biohealth, Wuchang Shouyi University, Wuhan 430064, P.R. China

² Institute of Crop, Wuhan Academy of Agricultural Sciences, Wuhan 430065, P.R. China

³ School of Horticulture and Landscape Architecture, Hubei Vocational College of Bio-Technology, Wuhan 430070, P.R. China

⁴ National Key Laboratory for Tropical Crop Breeding, Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, P.R. China

⁵ School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, P.R. China

⁶ Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya 572025, P.R. China

*Corresponding authors: E-mails: zhangna@wuhanagri.com, huangxing@catas.cn

Abstract

Background and aims: 4-Coumarate:CoA ligase (4CL) is a key enzyme in the phenylpropanoid pathway, directing metabolic flux toward lignin and flavonoid biosynthesis. This study aimed to identify and characterize the 4CL gene family (*Ah4CLs*) in *Agave* and to analyze their expression patterns during leaf development and under heavy metal and biotic stresses.

Methods: *Ah4CL* genes were identified from transcriptome data. Phylogenetic analysis was performed using the Maximum Likelihood (ML) method with IQ-TREE. Expression patterns were examined by qPCR across three leaf developmental stages and under copper (Cu), lead (Pb), and *Phytophthora nicotianae* treatments.

Results: Phylogenetic analysis classified the five genes into three groups: Group I (*Ah4CL1*, predicted lignin-related), Group II (*Ah4CL2*, predicted flavonoid-related), and Group III (*Ah4CL3-5*, 4CL-like). Expression profiling revealed that *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* were highly expressed in unexpanded leaves and were coordinately upregulated under Cu and Pb stress. Conversely, *Ah4CL1* and *Ah4CL2* were significantly induced by *P. nicotianae* infection.

Conclusions: The *Ah4CL* gene family exhibits functional diversification. *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* are primarily associated with leaf development and heavy metal tolerance, while *Ah4CL1* and *Ah4CL2* play a significant role in biotic defense. These findings provide a foundation for the genetic improvement of stress resilience in *Agave*.

Keywords: 4-coumarate:CoA ligase genes, *Agave*, gene expression, phylogeny, qPCR.

Introduction

Phenylpropanoid metabolism is a central pathway in plant secondary metabolism, directing carbon flow from primary

metabolism into the biosynthesis of lignin, flavonoids, phenolic acids, and other specialized metabolites. 4CL catalyzes the formation of hydroxycinnamoyl-CoA esters, serving as a key entry point into various downstream

Received 7 December 2025, last revision 13 August 2026, accepted 14 September 2026.

Abbreviations: 4CL - 4-coumarate:CoA ligase; ACS - acyl-CoA synthetases; A. H11648 - *Agave* hybrid 11648; CAD - cinnamyl alcohol dehydrogenase; CAM - crassulacean acid metabolism; CCR - cinnamoyl-CoA reductase; CDS - coding sequence; CesA - cellulose synthase; CHI - chalcone isomerase; CHS - chalcone synthase; hpi - hours post-infection; MW - molecular weight; ORF - open reading frame; PAL - phenylalanine ammonia-lyase; PP2A - protein phosphatase 2A; qPCR - real-time quantitative PCR; ROS - reactive oxygen species; TUB - tubulin.

Acknowledgements: This research was funded by Hainan Provincial Natural Science Foundation of China (323RC544), National Natural Science Foundation of China (32001598), China Agriculture Research System of MOF, MARA (CARS-16), and Central Public-Interest Scientific Institution Basal Research Fund (1630042022005, 1630042025008).

Data availability statement: All data are contained within the article or supplementary materials.

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

branches of the pathway (Lavhale *et al.*, 2018). Using ATP and Mg^{2+} as cofactors, 4CL converts several cinnamic acid derivatives (including cinnamic acid, 4-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid, and sinapic acid) into their corresponding acyl-CoA thioesters (Li *et al.*, 2020a). This enzyme modulates intracellular CoA ester levels through its action on different substrates and contributes to lignin biosynthesis *via* the activities of cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD). As a key rate-limiting enzyme in the synthesis of phytochemical monomers in vascular plants, 4CL has garnered significant attention in genetic engineering studies aimed at modulating lignin content (Liu *et al.*, 2018; Li *et al.*, 2020a). Research has shown that 4CL can regulate the expression of lignin biosynthesis-related genes, influence lignin accumulation, and thereby affect plant adaptability to drought and salt stress (Wang *et al.*, 2025; Zhang *et al.*, 2025c). Additionally, 4CL supports flavonoid production through its involvement with chalcone synthase (CHS) or chalcone isomerase (CHI) (Harding *et al.*, 2002).

Since the initial isolation of 4CL cDNAs from parsley (*Petroselinum hortense*) (Kuhn *et al.*, 1984; Lozoya *et al.*, 1988) and potato (*Solanum tuberosum*) (Becker-André *et al.*, 1991), the 4CL gene family has been widely investigated for its structural diversity, regulatory complexity, and biotechnological potential. In plants, 4CL constitutes a medium-sized gene family whose members have functionally diverged to participate in the biosynthesis of diverse phenylpropanoid derivatives, including lignin and flavonoids. Beyond their canonical roles, 4CL enzymes also participate in ubiquinone (coenzyme Q) biosynthesis. For example, in *Arabidopsis thaliana*, 4-CL8 (AT5G38120) localizes to peroxisomes and contributes to the formation of the benzenoid ring of coenzyme Q (Soubeyrand *et al.*, 2019). More critically, this superfamily contains a subclade of proteins closely related to 4CLs but possessing distinct catalytic activities, namely acyl-CoA synthetases (ACS). In *Arabidopsis*, AT1G62940 encodes a peroxisomal ACS involved in fatty acid metabolism. Other ACS isoforms, such as AT4G05160, AT5G63380, and AT1G20510, are also localized to peroxisomes, where they participate in fatty acid β -oxidation and jasmonic acid biosynthesis (Kienow *et al.*, 2008). These peroxisomal homologues form a separate phylogenetic clade from the *bona fide* 4CLs, despite sharing the conserved AMP-binding domain. 4CL proteins typically contain two conserved motifs: Box I (SSGTTGLPKGV), which corresponds to the AMP-binding domain, and Box II (GEICIRG), which is essential for catalytic activity (Costa *et al.*, 2005; Lavhale *et al.*, 2018). Plant 4CL genes are typically classified into two major phylogenetic clades: Class I (lignin-related) and Class II (non-lignin or flavonoid-related). Class I 4CLs, including *Arabidopsis At4CL1* (AT1G51680), *At4CL2* (AT3G21240), *At4CL3* (AT1G65060), and *At4CL4* (AT3G21230, also referred to as *At4CL5* in some literature) (Ehlting *et al.*, 1999) as well as rice *Os4CL1* (Q6ETN3), *Os4CL3* (P17814), *Os4CL4* (Q67W82), and *Os4CL5* (Q6ZAC1) are primarily involved in monolignol biosynthesis for lignin deposition and cell

wall reinforcement, with some members also contributing to soluble phenolic esters like sinapoylmalate (Affi *et al.*, 2022). Class II 4CLs, represented by *Arabidopsis At4CL3* (AT1G65060) and rice *Os4CL2* (Q6ZAC1), predominantly function in flavonoid metabolism, anthocyanin production, and UV stress responses, often exhibiting distinct tissue-specific expression patterns such as high expression in anthers and inducibility by UV radiation. Notably, some 4CL members display unique substrate specificities. For instance, *At4CL4* specifically activates sinapate, while *Os4CL4* participates in tricin biosynthesis and phosphorus acquisition, highlighting the functional diversification of this gene family across different plant lineages. Therefore, when conducting homology searches to identify 4CL genes, it is crucial to distinguish true 4CLs from ACS and 4CL-like sequences. In the present study, we employed a multi-step screening strategy using conserved motif analysis (Box I and Box II) and excluding sequences with high similarity to peroxisome-targeted ACS proteins to ensure the accurate identification of five Ah4CL candidates from the sisal transcriptome. Comparable classification schemes and functional divergence have also been reported in tobacco (*Nicotiana tabacum*) (Lee and Douglas, 1996), pine (*Pinus*) (Zhang and Chiang, 1997), apple (*Malus domestica*) (Ma *et al.*, 2023) and walnut (*Juglans regia*) (Ma *et al.*, 2024). Recent genome-wide studies in wheat (*Triticum aestivum*) and *Brassica* species have further expanded this classification framework. In wheat, 110 putative *Ta4CL* genes were identified and classified into seven clades, with *Ta4CL91* showing pronounced responsiveness to drought and salt stress (Zhang *et al.*, 2025c). Across *Brassica* U's triangle species, a total of 53 4CL genes were identified and phylogenetically grouped into four distinct clusters (Zhang *et al.*, 2025a). Expression analyses revealed that 4CL genes exhibit tissue-specific patterns and substantial variation in transcript levels across developmental stages, reflecting their functional diversification (Fu *et al.*, 2025). In apple, the expression of different 4CL gene members varied significantly depending on cultivar and tissue type, with generally higher levels observed in fruits, flowers, and leaves. Moreover, their expression was positively correlated with the intensity of fruit coloration and anthocyanin accumulation (Ma *et al.*, 2023).

Accumulating evidence underscores that 4CL genes are transcriptionally regulated in response to diverse abiotic stresses, playing a pivotal role in plant adaptation. This regulation is evident at the promoter level, as seen in cassava where *Me4CL32* possesses *cis*-elements responsive to various hormones (Ran *et al.*, 2024). Functional studies across species consistently demonstrate that 4CL activation is a central strategy for stress mitigation. For instance, in maize and wheat, the induction of *Zm4CL2* and *Ta4CL91* by drought or salt stress, respectively, is crucial for enhancing tolerance through lignin deposition and oxidative stress reduction (Wang *et al.*, 2025; Zhang *et al.*, 2025c). This mechanism extends to other stresses and metabolic outputs. Overexpression of 4CL genes in *Salvia miltiorrhiza* and tobacco redirects metabolic flux towards phenolic acids or flavonoids, concomitantly improving

growth and stress resilience (Li et al., 2020a). A detailed study on maize *Zm4CL2* further revealed that it coordinates drought-responsive lignin biosynthesis by interacting with *ZmCCoAOMT* (Wang et al., 2025). Notably, this paradigm of 4CL-mediated cell wall reinforcement is also a key defense against heavy metal toxicity. Under conditions like copper and lead stress, 4CL upregulation promotes lignin biosynthesis, which reinforces the cell wall and sequesters metal ions, thereby alleviating oxidative damage and cellular injury (Le Gall et al., 2015; Tugbaeva et al., 2022). Therefore, the transcriptional regulation of 4CL genes represents a conserved regulatory node that orchestrates the biosynthesis of protective phenylpropanoids, enabling plants to cope with a wide spectrum of environmental challenges.

To date, no systematic study has been conducted to identify members of the 4CL gene family in *Agave* or to investigate their evolutionary relationships and expression patterns. Transcriptome sequencing has emerged as a powerful tool for gene discovery in non-model plants such as *Agave*, whose large and complex genomes have hindered whole-genome sequencing efforts. Huang et al. performed *de novo* transcriptome assembly of *Agave* hybrid 11648 (*A. H11648*), generating 148 046 unigenes and identifying 38 *CesA* genes that are critical for cellulose biosynthesis during fiber formation (Huang et al., 2019). Wang et al. sequenced the transcriptome of *A. amaniensis*, identifying nine *expansin A* genes involved in cell wall loosening and fiber elongation (Wang et al., 2023). Most *expansin* genes exhibited high expression in shoot tissues, with expression levels declining during leaf maturation, suggesting a negative correlation with leaf elongation. Notably, *AhEXP2* was significantly upregulated under biotic stress induced by *Phytophthora nicotianae* infection, indicating its potential role in disease resistance. Beyond these studies, Wang et al. reported the first transcriptome of *A. schidigera*, identifying 72 160 unigenes and cloned 20 *PAL* genes from four *Agave* species and analyzed their expression under different developmental and stress conditions (Wang et al., 2024). Furthermore, our team successfully applied *PacBio* sequencing to obtain full-length transcriptomes from five distinct tissues of *Agave* leaves (Huang et al., 2024). This dataset represents the first comprehensive full-length transcriptomic resource for *Agave*, providing a valuable foundation for gene discovery and functional studies in this genus.

Agave species are economically important CAM plants with diverse agricultural applications, including tequila production and natural fiber manufacturing. *A. H11648* is the predominant cultivated variety for fiber production in tropical regions worldwide. Therefore, investigating the 4CL gene family in sisal may not only facilitate the accumulation of valuable metabolites such as flavonoids and lignin but also improve the plant's resilience to environmental stresses. Furthermore, understanding the evolution and function of these genes can offer valuable insights into the genetic basis of sisal development. Based on the available transcriptome data for *A. H11648*, this study identified and cloned the 4CL genes, systematically analyzed their phylogenetic relationships and expression

patterns during leaf development and characterized their differential expression under copper stress, lead stress, and pathogen infection. These findings will enhance our understanding of the functional roles of 4CL genes in sisal fiber development and in stress and disease resistance mechanisms, thereby providing a theoretical foundation for breeding high-yielding, high-quality, and multi-resistant sisal varieties. In this study, we classified both 4CL and 4CL-like genes into the 4CL gene family and designated them collectively under the *Ah4CL* nomenclature.

Materials and methods

Material cultivation and processing: The main *Agave* cultivar *A. H11648* was cultivated at the Institute of Environment and Plant Protection, Chinese Academy of Tropical Agricultural Sciences (19.99°N, 110.33°E). Two-year-old potted plants were sampled at three leaf developmental stages: heart leaf, unexpanded leaf, and fully expanded leaf (defined as having an angle greater than 45° from the central axis). One-year-old potted plants were used for stress treatments. Heavy metal stress was applied by irrigating with aqueous solutions of copper sulfate or lead nitrate, using clean water irrigation as the control. Based on established protocols, two concentrations were used: 1 g/L CuSO₄ and 1.3 g/L Pb(NO₃)₂. Each treatment received 500 mL of the respective solution, and leaf samples were collected approximately two weeks post-treatment when curling symptoms appeared due to heavy metal toxicity. For pathogen infection, detached leaves were wound-inoculated with *P. nicotianae*, the causal agent of zebra spot disease in *Agave*. Control leaves were wounded but not inoculated. Sampling was performed at 24 and 48 h post-inoculation. For each developmental stage, stress treatment, and control, samples were collected from three independent plants as biological replicates. All collected tissues were immediately flash-frozen in liquid nitrogen and stored at -80°C until use. Total RNA was extracted using a *Plant Total RNA Extraction Kit* (Tiangen Biotech, Beijing, China) and preserved for subsequent analysis.

Identification and analysis of the *Ah4CL* genes:

To ensure the identification of *bona fide* 4CL genes rather than ACS or other 4CL-like homologs, we employed a multi-step screening strategy. First, using functionally characterized 4CL sequences from *A. thaliana* (*At4CL1* - *At4CL4*) and rice (*Os4CL1* - *Os4CL5*) (Hamberger and Hahlbrock, 2004; Gui et al., 2011) as queries, we performed a systematic search of the *Agave* transcriptome database using the *TBlastx* algorithm. Only sequences showing an *E*-value of $\leq 1 \times 10^{-10}$ in the BLAST results were retained as candidate members of the *Ah4CL* gene family. The identified transcripts were subsequently analyzed for ORFs using the *ORF-FINDER* tool, and only those containing a complete coding sequence were retained for further analysis. Second, candidate sequences were examined for the presence of complete and conserved catalytic motifs, including the AMP-binding domain (Box I: SSGTTGLPKG_V) and the putative

catalytic site (Box II: GEICIRG). For multiple sequence alignment of the identified *Agave* 4CL protein sequences, we used *DNAMAN* software (Lynnon Biosoft, San Ramon, CA, USA) to identify conserved amino acid sequences within functional domains. The protein length, molecular weight, and pI were predicted using the *ProtParam* tool (*ExPASy*, SIB, Geneva, Switzerland), while subcellular localization was predicted with the *CELLO* prediction system (National Taiwan University & Academia Sinica, Taipei, Taiwan, China). Third, sequences showing higher similarity to peroxisome-targeted ACS proteins were excluded from further analysis. This rigorous approach allowed us to identify five sequences (*Ah4CL1* - *Ah4CL5*) that most likely represent authentic 4CL enzymes in *Agave*. To construct a phylogenetic tree, we selected the gene sequences of *A. thaliana* (Cukovica *et al.*, 2001), *Oryza sativa* (Gui *et al.*, 2011), *Asparagus officinalis* (Harkess *et al.*, 2017), *Brassica oleracea* (Zhang *et al.*, 2025a), *Solanum tuberosum* (Nie *et al.*, 2023), and *Zea mays* (Zhang *et al.*, 2025b) as reference sequences. Multiple sequence alignment was performed using *MAFFT* with the *L-INS-i* algorithm (Katoh and Standley, 2013). Phylogenetic analysis was conducted using the Maximum Likelihood (ML) method with *IQ-TREE* (Minh *et al.*, 2020). The best-fit amino acid substitution model was selected by *ModelFinder* (Kalyaanamoorthy *et al.*, 2017). Branch supports were evaluated by ultrafast bootstrap (1 000 replicates) and *SH-aLRT* test (1 000 replicates).

Expression patterns of the *Ah4CL* genes: Transcriptome expression data were obtained from the previously established *de novo* transcriptome assembly of *A. H11648*, which was generated by our team through *Illumina* sequencing (Huang *et al.*, 2019). Total RNA was extracted from *Agave* samples and reverse-transcribed into cDNA using the *GoScript™ Reverse Transcription System* (Promega, Madison, WI, USA) for subsequent qPCR analysis. The qPCR amplifications were performed on a *QuantStudio™ 6 Real-Time PCR System* (Thermo Fisher, Waltham, MA, USA) with a total reaction volume of 20 µL. Each reaction mixture contained 1 µL of cDNA template, 10 µL of *TransStart® Tip Green qPCR Supermix* (Beijing Quanshijin Biotechnology Co., Ltd., Beijing, China), 0.5 µL of forward primer (10 µM), 0.5 µL of reverse primer (10 µM), 0.4 µL of *Passive Reference Dye* (Beijing Quanshijin Biotechnology Co., Ltd., Beijing, China), and 7.6 µL of nuclease-free water. The thermal

cycling protocol consisted of an initial denaturation at 94°C for 3 min, followed by 40 cycles of denaturation at 94°C for 10 s and annealing at 58°C for 30 s, and concluded with a melting curve analysis. Each sample was analyzed in three technical replicates.

The *PP2A* and *TUB* genes were used as reference genes to normalize the expression of target genes. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Specific qPCR primers for the reference genes and the *4CL* genes were designed using *Primer3* software (Untergasser *et al.*, 2012), and their sequences are listed in Table 1.

Results

Identification and cloning of *Ah4CL* family: Through homologous sequence analysis, five full-length coding sequences were identified and designated as *Ah4CL1*, *Ah4CL2*, *Ah4CL3*, *Ah4CL4*, and *Ah4CL5*. These sequences have been submitted to the *GenBank* database. To distinguish *bona fide* 4CL enzymes from 4CL-like homologs that may possess ACS activity, a systematic comparative analysis of conserved domains was performed. Comparative analysis of conserved domains revealed that *Ah4CL1* and *Ah4CL2* contain the characteristic AMP-binding motifs SSGTTGLPKGV (Box I) and GEICIRG (Box II), consistent with previously reported 4CL sequences (Ehlting *et al.*, 2001; Schneider *et al.*, 2003; Zhang *et al.*, 2025a). Box I and Box II are highly conserved core domains within the adenylate-forming enzyme superfamily. Box I constitutes the AMP-binding site and directly participates in the catalytic process, while the integrity of Box II is essential for 4CL enzyme activity-deletion or mutation of this motif results in near-complete loss of catalytic function (Ma *et al.*, 2026). In contrast, Box I of *Ah4CL4* and *Ah4CL5* each contained two amino acid substitutions, where the conserved LP residues were replaced by VS and DS, respectively. Box I of *Ah4CL3* (SSGTS GFSGKV) showed three amino acid variations: T to S, and LP to FS. Similarly, Box II regions in *Ah4CL4*, *Ah4CL5*, and *Ah4CL3* displayed variant sequences GELLFRG, GELCFRG, and GELFLRG, respectively (Fig. 1). The SGT region represents a serine/threonine/glycine-rich segment, immediately followed by a conserved proline/serine-lysine-glycine triplet. These evolutionarily conserved sequences are essential for

Table 1. Primers used for qPCR analysis.

Genes	Forward primer (5'- 3')	Reverse primer (5'- 3')	Product size (bp)
<i>Ah4CL1</i>	CAAACGGGGCTGAAATCACA	GGTCGGATGGAGATCACTGT	198
<i>Ah4CL2</i>	ACAAGTCCCTGGTATCGAGC	ATCACCTCCAGCATCCTCTC	205
<i>Ah4CL3</i>	TCTTGACGCTCTACCGATGT	ATCCGTTTCGCATGTTTAGCC	203
<i>Ah4CL4</i>	CTCCTGTACTCTCTGGCAC	GAACCTCTCCATCACGACCA	213
<i>Ah4CL5</i>	TGCTGAACTTGAAGGGTTGC	GATGCGGATTTTGGGATGCT	225
<i>PP2A</i>	CCTCCTCCTCCTTCGGTTTG	GCCATGAATGTCACCGCAGA	235
<i>TUB</i>	TTCCCATCACAAAGGTCTC	CGCTCATTGTGGCAGAGATA	196

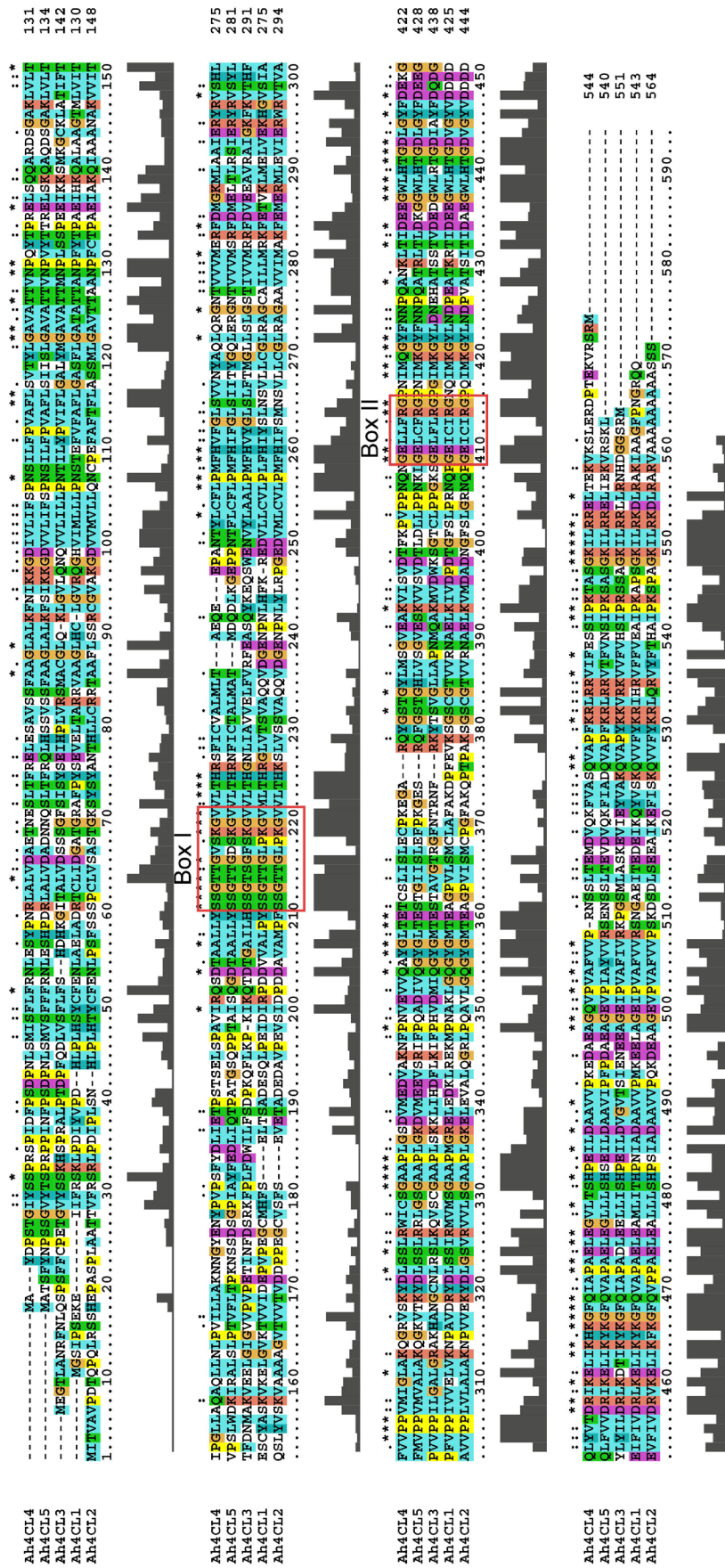


Fig. 1. Multiple amino acid sequence alignment of the Ah4CL domains in *A. H11648*. The AMP-binding enzyme domains were aligned. The red boxes indicate the conserved peptide motifs **Box I** and **Box II**.

maintaining the functional integrity of the AMP-binding domain (Schróder, 1989; Nie *et al.*, 2023). Additionally, valine (V) was found to be highly conserved within the AMP-binding domain across all *Agave* 4CL proteins.

The CDS length, amino acid composition, MW, pI, and subcellular localization of each member of the Ah4CL family were systematically analyzed. The CDS range from 1 623 to 1 695 bp, encoding proteins of 540 - 564 amino acids with predicted molecular weights of approximately 60 kDa (59.1 - 60.8 kDa) and theoretical isoelectric points between 5.4 and 9.0 (Table 2). Subcellular localization predictions indicate that Ah4CL1 is localized in the cytoplasm, Ah4CL2 in the chloroplast, and Ah4CL3 - Ah4CL5 are predicted to localize to both chloroplast and plasma membranes.

Phylogenetic analysis of *Ah4CL* family: To further elucidate the evolutionary relationships and infer the potential functions of the five *Ah4CL* genes, a phylogenetic tree was constructed, including 4CL homologs from multiple plant species (Fig. 2). The results showed that the five *Ah4CL* genes were divided into three distinct clusters (Groups I - III).

Group I contained *Ah4CL1*, which clustered with *At4CL1*, *At4CL2*, *At4CL4* from *A. thaliana*, *Os4CL1*, *Os4CL3*, *Os4CL4*, *Os4CL5* from *O. sativa*, as well as multiple Class I 4CL genes from maize (*Z. mays*), cabbage (*B. oleracea*) and potato (*S. tuberosum*). Previous studies have demonstrated that these genes are primarily involved in lignin biosynthesis (Cukovica *et al.*, 2001; Gui *et al.*, 2011). Therefore, based on phylogenetic inference,

Table 2. Physicochemical properties and subcellular localization of 4CL genes and their encoded proteins.

Genes	Accession	CDS length (bp)	Protein length (aa)	MW (kDa)	Theoretical pI	Subcellular localization
<i>Ah4CL1</i>	transcript50796/f41p0/2002	1 632	543	59.6	5.6	cytoplasmic
<i>Ah4CL2</i>	transcript46176/f3p0/2091	1 695	564	59.8	5.4	chloroplast
<i>Ah4CL3</i>	transcript44843/f2p0/2202	1 656	551	60.8	9.0	chloroplast; plasma membrane
<i>Ah4CL4</i>	transcript54830/f23p0/1883	1 635	544	59.9	5.9	chloroplast; plasma membrane
<i>Ah4CL5</i>	transcript52736/f3p0/2024	1 623	540	59.1	7.1	chloroplast; plasma membrane

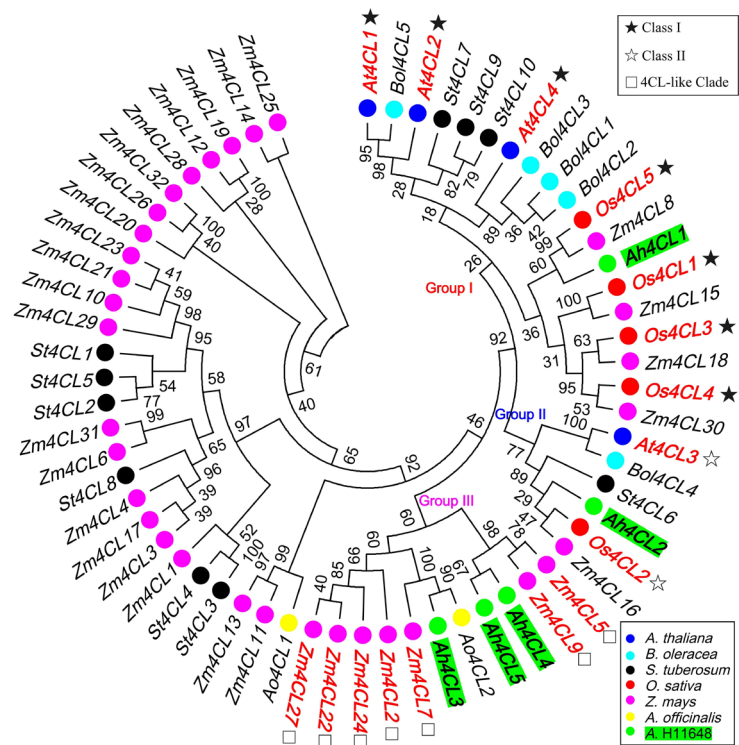


Fig. 2. Phylogenetic relationships of the 4CL genes family. Multiple sequence alignment by MAFFT (*L-INS-i* algorithm). Phylogenetic tree constructed using IQ-TREE with the Maximum Likelihood method; optimal substitution model selected by ModelFinder; branch support assessed by ultrafast bootstrap (1 000 replicates) and SH-aLRT test (1 000 replicates).

Ah4CL1 was assigned to Group I, which corresponds to the previously established Class I 4CL clade, suggesting that it may contribute to lignin accumulation in *Agave*, pending direct functional validation. Group II contained *Ah4CL2*, which clustered with *At4CL3*, *Os4CL2*, *Zm4CL16*, *St4CL6*, and *Bol4CL4*. *At4CL3* and *Os4CL2* have been confirmed to participate in flavonoid biosynthesis and belong to Class II 4CL. Accordingly, *Ah4CL2* was assigned to Group II, corresponding to the previously established Class II 4CL clade, suggesting its potential involvement in flavonoid metabolism, which requires further experimental confirmation. Group III represents a 4CL-like lineage rather than a canonical 4CL Class I or Class II. Group III contained three genes (*Ah4CL3*, *Ah4CL4*, and *Ah4CL5*), which clustered with maize *Zm4CL9*, *Zm4CL5*, *Zm4CL2*, *Zm4CL7*, *Zm4CL22*, *Zm4CL24*, *Zm4CL27*, and asparagus *Ao4CL2*. Notably, these maize 4CL genes all belong to the 4CL-like Clades (Clade A - E), rather than Clade 4CL, which is involved in the core biosynthesis of lignin and flavonoids (Zhang et al., 2025b). Therefore, *Ah4CL3*, *Ah4CL4*, and *Ah4CL5* were classified as 4CL-like genes. Their biological functions remain to be determined, as they may differ significantly from those of canonical 4CL genes. On the phylogenetic tree, these three *Agave* genes formed a species-specific branch, suggesting that they may have undergone functional diversification in *Agave*. Additionally, genes asparagus *Ao4CL1*, 4CL gene series from maize (*Z. mays*) and potato (*S. tuberosum*) clustered into several distinct clades, which contained no *Agave* 4CL genes.

Expression patterns of *Ah4CL* family: To examine the expression patterns of *Ah4CL* genes, we performed qPCR analysis on five genes (*Ah4CL1* - *Ah4CL5*) at different leaf developmental stages in *A. H11648* (Fig. 3). *Ah4CL1* and *Ah4CL4* showed high expression levels in the shoot, unexpanded leaf, and expanded leaf, with particularly strong expression in unexpanded leaves. *Ah4CL5* displayed a similar expression pattern to *Ah4CL1* and *Ah4CL4*, though its transcript levels were consistently lower across all stages. *Ah4CL2* was also highly expressed

in shoots, but its expression gradually decreased during leaf development, reaching the lowest level in fully expanded leaves. In contrast, *Ah4CL3* exhibited consistently low expression throughout all three growth stages, distinguishing it from the other four *Ah4CL* genes.

To investigate the potential involvement of *Ah4CL* genes in abiotic stress responses, we analyzed their expression patterns in *Agave* under copper (Cu) and lead (Pb) treatments. The results showed that *Agave* exhibits significant tolerance to Cu and Pb stress. *Agave* exhibited a significant tolerance phenotype, characterized by reduced growth but maintained tissue integrity without rapid necrosis under both metal stresses. Expression profiling revealed that all five *Ah4CL* genes were upregulated under Cu stress. Notably, *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* exhibited high constitutive expression levels. Similarly, under Pb stress, the expression of *Ah4CL1*, *Ah4CL2*, *Ah4CL4*, and *Ah4CL5* increased compared to the control, with *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* again showing consistently high expression in both control and treated groups (Fig. 4). These results indicate that *Ah4CL* genes, particularly *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* are responsive to heavy metal stress in *Agave*.

During their growth and development, sisal plants are highly susceptible to pathogenic microbial infections with *P. nicotianae* Breda being the most prevalent causal agent of zebra spot disease. To investigate the response of *Ah4CL* genes to biotic stress, we analyzed their expression patterns in *Agave* leaves at 0, 24, and 48 h following *P. nicotianae* inoculation. The results showed that *Ah4CL1* and *Ah4CL2* were progressively induced during infection, with transcript levels increasing approximately threefold at 48 hpi compared to the control (T0). *Ah4CL4* exhibited relatively high basal expression in uninfected leaves, which was further elevated nearly twofold at 24 hpi but subsequently declined to T0 levels by 48 hpi. In contrast, *Ah4CL5* maintained consistently high expression across all time points without significant alteration upon pathogen challenge. Likewise, *Ah4CL3* showed no significant induction and remained at low expression levels throughout the infection course (Fig. 5).

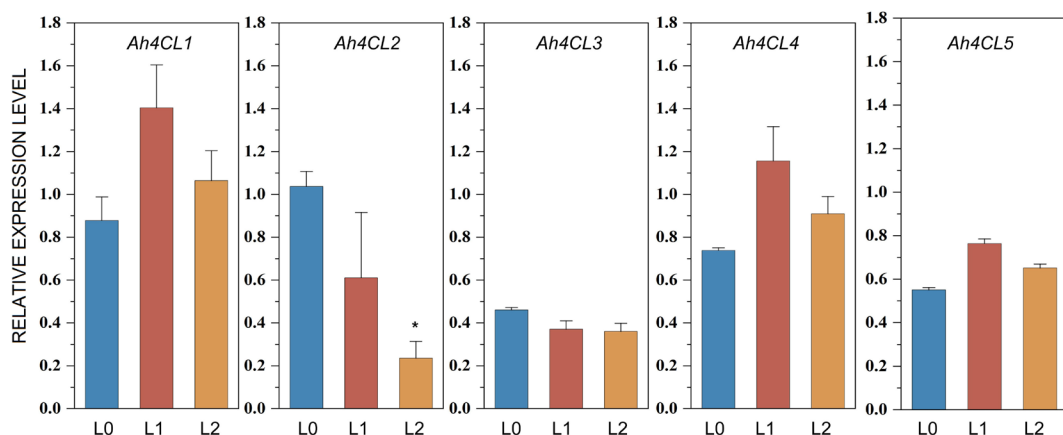


Fig. 3. Expression pattern of *Ah4CL* genes at three different leaf developmental stages according to qPCR. Data are presented as mean $\pm$ SE ($n = 3$). The gene with expression level decreased more than 3-fold was marked with the asterisk. L0 - shoot, L1 - unexpanded leaf, L2 - expanded leaf.

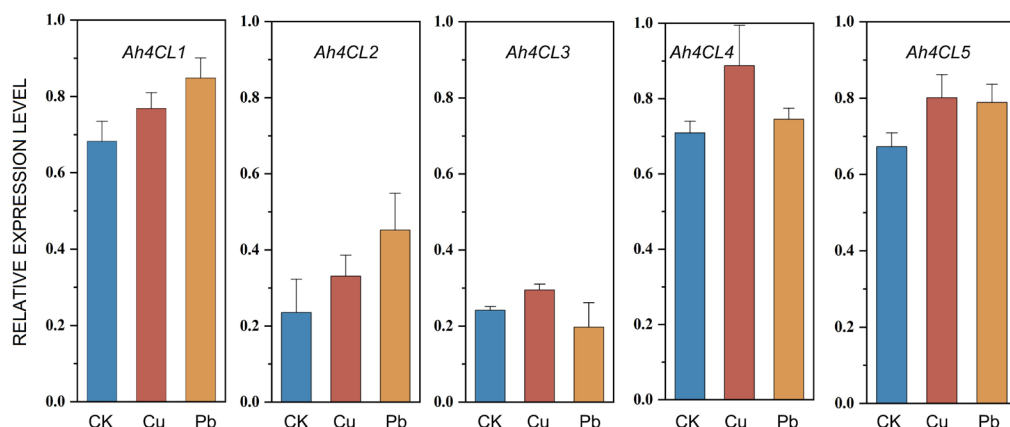


Fig. 4. Expression pattern of *Ah4CL* genes under abiotic stresses, including Cu and Pb treatments determined by qPCR. Data are presented as mean $\pm$ SE ($n = 3$). CK - control.

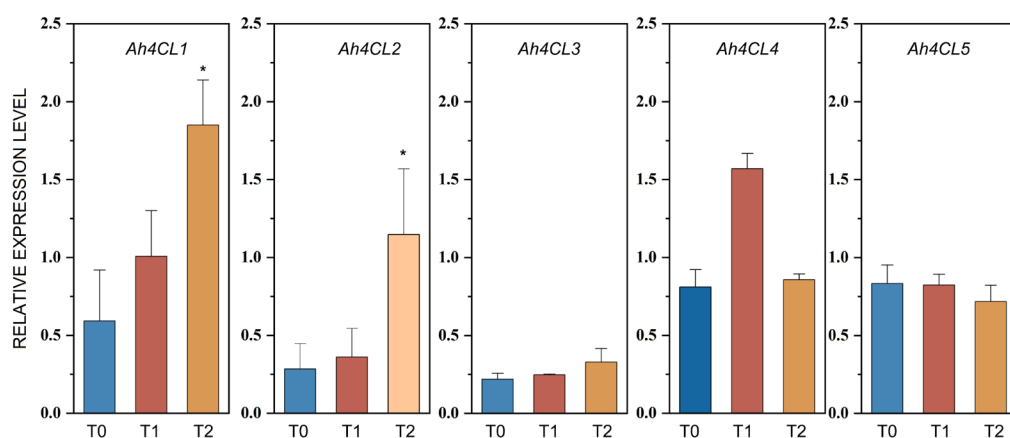


Fig. 5. Relative expression of *Ah4CL* genes under biotic stress (*P. nicotianae* inoculation) determined by qPCR. Data are presented as mean $\pm$ SE ($n = 3$). T0 - control, T1 - 24 hpi, T2 - 48 hpi. Genes with more than 3-fold increase of their expression are marked with the asterisk.

Discussion

4CL has a crucial role in the biosynthesis of plant secondary metabolites such as flavonoids and lignin where general phenylpropanoid pathways diverge into multiple major metabolic routes (Hamberger and Hahlbrock, 2004). 4CL catalyzes the corresponding cinnamic acid and its hydroxyl or methoxy derivatives to generate the corresponding coenzyme A ester. The intermediate products catalyzed by different 4CLs are respectively distributed to the metabolic branches of phenylpropane derivatives such as flavonoids or lignin for the synthesis of flavonoids or lignin. Therefore, cloning the 4CL gene and clarifying its function in the fiber development process of *Agave* and the response to biological and abiotic stresses are of great significance for promoting the genetic improvement and breeding of *Agave*. Due to the large size and high heterozygosity of the *Agave* genome, its genome assembly faces significant challenges (Wang *et al.*, 2023). In recent years, advances in transcriptome sequencing have provided an efficient, high-throughput, and cost-effective platform for functional gene studies in *Agave*. Based on the published transcriptome data from *Agave* leaves (Huang *et al.*, 2019), we identified

and sequenced five *Ah4CL* genes from the transcriptome of *A. H11648*, encoding proteins ranging from 540 to 564 amino acids. This represents the first report of these genes in *Agave*. Subcellular localization predictions indicated that *Ah4CL1* localized to the cytoplasmic, *Ah4CL2* localized to the chloroplast, and *Ah4CL3* - *Ah4CL5* to the plasma membrane and chloroplast. Such compartmentalization may facilitate the channeling of substrates toward distinct metabolic pathways, consistent with observations in other plant species where 4CL isoforms are distributed across multiple organelles (Nie *et al.*, 2023).

On the phylogenetic tree, *Ah4CL1* and *Ah4CL2* were classified into Class I and Class II, respectively. However, it is important to note that these classifications are derived from sequence similarity and phylogenetic clustering with functionally characterized homologs in other species, not from direct functional evidence in *Agave*. The clustering of *Ah4CL3*, *Ah4CL4*, and *Ah4CL5* is more distinctive. These three genes clustered together with maize *Zm4CL9*, *Zm4CL5*, *Zm4CL2*, *Zm4CL7*, *Zm4CL22*, *Zm4CL24*, *Zm4CL27*, and *Ao4CL2* into the same subgroup. According to the classification of the maize 4CL gene

family, these maize genes all belong to the *At4CL-like* clades (Clade A - E). Therefore, *Ah4CL3-5* should be classified as *4CL-like* genes, and their functions may differ significantly from those of canonical *4CL* genes. In maize, *Zm4CL2* and *Zm4CL7* from the same subgroup were significantly upregulated under UV treatment, while *Zm4CL27* showed specific high expression in anthers (Zhang et al., 2025b), suggesting that this type of gene may participate in photooxidative stress responses or reproductive development. In potato, several *4CL-like* genes have also been reported to be differentially expressed under various stress conditions (Nie et al., 2023). These findings suggest that *Ah4CL3-5* may not directly contribute to lignin or flavonoid biosynthesis, but rather play roles in stress adaptation or other secondary metabolic pathways in *Agave*.

The conserved AMP-binding domain (SSGTTGLPKGV) and Box II motif (GEICIRG) were identified in *Ah4CL1* and *Ah4CL2*, whereas *Ah4CL3*, *Ah4CL4*, and *Ah4CL5* contained variations at specific sites. Based on functional studies of *4CL* genes in rice (Gui et al., 2011; Sun et al., 2013) and *A. thaliana* (Ehlting et al., 1999; Hamberger and Hahlbrock, 2004), along with the cluster analysis results, *Ah4CL1* was predicted to participate in lignin biosynthesis and *Ah4CL2* predicted to be involved in the flavonoid biosynthetic pathway, while the biological functions of *Ah4CL3* to *Ah4CL5* require further experimental investigation and functional validation. These two motifs are highly conserved across angiosperms, monocotyledons, and gymnosperms, constituting the core structural elements essential for the catalytic activity of the *4CL* proteins (Schneider et al., 2003; Wang et al., 2022; Zhang et al., 2025a). The presence of the highly conserved AMP-binding domain has served as the primary criterion for classifying enzymes including *4CLs*, firefly luciferases, acetyl-CoA synthetases, fatty acyl-CoA synthetases, and nonribosomal peptide synthetases into a single superfamily of adenylate-forming enzymes (Fulda et al., 1994). However, there are also literature reports indicating that mutations do occur at specific sites within these conserved sequences. For instance, in *Dryopteris fragrans*, Box I of *Df4CL3* and *Df4CL4* were SSGTTGVSKGV, with two divergent amino acids (LP to VS), while their Box II motifs display as variant GEIWFRG, containing two altered amino acids (CI to WF) (Li et al., 2015). Similarly, multiple sequence alignment of conserved domains in *Camellia sinensis* *4CL* genes showed that five genes contained the conserved Box I (SSGTTGLPKGV) and Box II (GEICIRG) motifs, while the remaining twelve genes exhibited mutations at two amino acid sites within these regions (Ren et al., 2023). The observed variations in the Box I and Box II conserved domains suggest that *4CL* genes may have undergone functional diversification during evolution, potentially leading to differences in substrate recognition, binding specificity, or catalytic activity.

The differential expression patterns of *Ah4CL* genes across developmental stages and their consistent upregulation under heavy metal stress provide important insights into their functional diversification in *Agave*.

The high constitutive expression of *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* in shoots and unexpanded leaves, as well as their further induction under stress, suggests their fundamental roles in both primary development and stress adaptation. These isoforms are likely central to the biosynthesis of cell wall components necessary for fiber formation and structural integrity, a premise supported by reports of specific *4CL* isoforms being highly expressed in lignin-rich tissues in other species such as *Solanum tuberosum* and *Salvia miltiorrhiza* (Nie et al., 2023). In many plant species, the tissue-specific expression patterns of *4CL* genes are closely associated with their functional divergence. For example, in the tea plant (*Camellia sinensis*), the Class I member *Cs4CL1* is predominantly expressed in roots and stems and is involved in lignin biosynthesis, whereas the Class II member *Cs4CL2* is predominantly expressed in young leaves and is involved in flavonoid biosynthesis (Li et al., 2022). Although we did not examine root and stem tissues, based on phylogenetic analysis, *Ah4CL1* clusters within the Class I clade, suggesting its potential function in lignified tissues such as roots and stems; conversely, *Ah4CL2* clusters within the Class II clade, suggesting its potential involvement in flavonoid biosynthesis in leaves. Future studies will include additional tissues (including roots and stems) to comprehensively characterize the tissue-specific expression patterns of *Ah4CL* genes and to validate their functional divergence.

Under copper and lead stress treatments, all five *Ah4CL* genes were upregulated, with *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* showing particularly strong induction. This suggested that *4CL* genes may contribute to heavy metal stress tolerance, as well as to drought and salt stress adaptation, potentially through mechanisms such as lignin deposition for cell wall reinforcement or the synthesis of protective flavonoids (Wang et al., 2025; Zhang et al., 2025c). For instance, in *Fraxinus mandshurica*, Cd stress specifically induces the expression of *Fm4CL1*. Its overexpression enhances Cd tolerance in transgenic plants by promoting lignin accumulation and boosting antioxidant capacity (Chen et al., 2019). Similarly, various heavy metal stresses have been widely reported to activate the phenylpropanoid metabolic pathway, leading to enhanced lignification and flavonoid synthesis (Le Gall et al., 2015). Based on these lines of evidence, we hypothesize that the upregulation of *Ah4CL* represents a core adaptive strategy of *Agave* in response to heavy metal toxicity. It may act through two synergistic mechanisms: (1) promoting lignin biosynthesis to reinforce the cell wall and sequestering Pb^{2+}/Cu^{2+} within the cell wall via chelation, thereby restricting their translocation into the cytoplasm; and (2) increasing the production of phenolic compounds such as flavonoids to scavenge the excessive ROS generated under heavy metal stress and mitigate oxidative damage. Thus, the induced expression of *Ah4CL* genes is likely a key component of a comprehensive, cell wall-mediated heavy metal detoxification mechanism in *Agave*. It is important to note that *4CL*, as a key enzyme in the phenylpropanoid pathway, functions not only in heavy metal stress responses but also plays critical roles

under various environmental stresses. Several studies have demonstrated that *4CL* genes are induced by drought, salt stress, mechanical wounding, and ultraviolet (UV) radiation, and enhance stress tolerance by regulating the accumulation of secondary metabolites such as lignin, flavonoids, and anthocyanins (Wang *et al.*, 2025; Zhang *et al.*, 2025c). Of particular relevance to *Agave*, anthocyanins have been reported as important metabolites in response to UV stress. For instance, anthocyanin (betacyanin) deposition in *Agave striata* leaves has been linked to photoprotective mechanisms, with red-leaf morphs accumulating higher anthocyanin content under field conditions (Fondom *et al.*, 2009), suggesting that UV stress may influence secondary metabolism in *Agave* through the *4CL*-mediated phenylpropanoid pathway. Furthermore, mechanical wounding and drought stress can induce *4CL* expression via activation of the jasmonic acid signaling pathway, promoting lignin and suberin deposition at wound sites or in vascular tissues to serve as physical barriers against further damage (Liu *et al.*, 2023).

Numerous studies have confirmed that *4CL* genes are induced in response to pathogen infection. In cotton, overexpression of *Gh4CL3* enhanced lignin and flavonoids accumulation conferring systemic resistance to *Verticillium dahliae* (Alariqi *et al.*, 2023), whereas ectopic expression of *Os4CL3* or *Os4CL5* in rice restricted *Magnaporthe oryzae* penetration through enhancing early lignin deposition (Li *et al.*, 2020b). In *Gastrodia elata*, two members *Ge4CL2* and *Ge4CL5* were significantly upregulated in tubers following fungal pathogen challenge (Sha *et al.*, 2025). Consistent with these reports, our study demonstrated that *Ah4CL1*, *Ah4CL2*, and *Ah4CL4* were significantly upregulated following *P. nicotianae* inoculation, with distinct temporal expression patterns. *Ah4CL1* and *Ah4CL2* showed progressive induction, while *Ah4CL4* displayed transient upregulation. These results support crucial regulatory roles for these *4CL* genes play in plant defense mechanisms. In contrast, *Ah4CL5* maintained consistently high expression regardless of pathogen challenge, suggesting constitutive expression for basal defense or involvement in alternative stress responses. *Phytophthora* species infection, including *P. nicotianae*, induces phenylpropanoid pathway genes and promotes flavonoid and lignin accumulation to enhance disease resistance. In strawberry, *4CL* genes were co-upregulated with flavonoid biosynthesis genes following *P. cactorum* inoculation (Toljamo *et al.*, 2016). Therefore, the differential upregulation of *Ah4CL* genes following *P. nicotianae* inoculation likely contributes to *Agave* defense through coordinated synthesis of antimicrobial flavonoids and lignin, with specific family members responding dynamically to infection timing while others provide sustained basal protection.

Conclusion and perspective

This study presents the first systematic identification and expression analysis of the *Ah4CL* gene family in *Agave*. Phylogenetic analysis classified the five *Ah4CL* genes into

three distinct groups: Group I (*Ah4CL1*, predicted lignin-related Class I), Group II (*Ah4CL2*, predicted flavonoid-related Class II), and Group III (*Ah4CL3*, *Ah4CL4*, and *Ah4CL5*, *4CL*-like genes with undetermined function). Expression analyses suggest that *Ah4CL1*, *Ah4CL4*, and *Ah4CL5* may be coordinately upregulated during leaf development and under heavy metal stress, consistent with potential roles in fiber formation and stress adaptation. Similarly, *Ah4CL1* and *Ah4CL2* were significantly induced by *P. nicotianae*, indicating their involvement in biotic stress responses. These findings provide a preliminary foundation for the genetic improvement of stress resilience in *Agave*. Future studies employing gene knockout, overexpression, and biochemical assays are essential to validate the precise functions of each *Ah4CL* member. With the advancement of *Agave* whole-genome sequencing, genome-wide identification of *4CL* family members and their regulatory networks will become feasible.

References

- Affi, O.A., Tobimatsu, Y., Lam, P.Y. *et al.* (2022) Genome-edited rice deficient in two *4-COUMARATE:COENZYME A LIGASE* genes displays diverse lignin alterations. *Plant Physiology*, 190, 2155-2172.
- Alariqi, M., Ramadan, M., Wang, Q. *et al.* (2023) Cotton 4-coumarate-CoA ligase 3 enhanced plant resistance to *Verticillium dahliae* by promoting jasmonic acid signaling-mediated vascular lignification and metabolic flux. *The Plant Journal*, 115, 190-204.
- Becker-André, M., Schulze-Lefert, P., Hahlbrock, K. (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato. *The Journal of Biological Chemistry*, 266, 8551-8559.
- Chen, X., Wang, H., Li, X., Ma, K., Zhan, Y. & Zeng, F. (2019) Molecular cloning and functional analysis of 4-Coumarate:CoA ligase 4 (*4CL-like 1*) from *Fraxinus mandshurica* and its role in abiotic stress tolerance and cell wall synthesis. *BMC Plant Biology*, 19, 231.
- Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A. *et al.* (2005) Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry*, 66, 2072-2091.
- Cukovica, D., Ehrling, J., Van Ziffle, J.A. & Douglas, C.J. (2001) Structure and evolution of 4-coumarate:coenzyme A ligase (*4CL*) gene families. *Biological Chemistry*, 382, 645-654.
- Ehrling, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E. & Kombrink, E. (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *The Plant Journal*, 19, 9-20.
- Ehrling, J., Shin, J.J.K. & Douglas, C.J. (2001) Identification of 4-coumarate:coenzyme a ligase (*4CL*) substrate recognition domains. *The Plant Journal*, 27, 455-465.
- Fondom, N.Y., Castro-Nava, S. & Huerta, A.J. (2009) Photoprotective mechanisms during leaf ontogeny: cuticular development and anthocyanin deposition in two morphs of *Agave striata* that differ in leaf coloration. *Botany*, 87, 1186-1197.
- Fu, F., Qin, H., Xin, Y. *et al.* (2025) Characterization of 4-coumarate CoA ligase (*4CL*) gene family and functional study of Sm4CL2/3/7/9 in *Salvia miltiorrhiza*. *Functional &*

- Integrative Genomics*, 25, 35.
- Fulda, M., Heinz, E. & Wolter, F.P. (1994) The *fadD* gene of *Escherichia coli* K12 is located close to *rnd* at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family. *Molecular & General Genetics*, 242, 241-249.
- Gui, J., Shen, J. & Li, L. (2011) Functional characterization of evolutionarily divergent 4-coumarate:coenzyme A ligases in rice. *Plant Physiology*, 157, 574-586.
- Hamberger, B. & Hahlbrock, K. (2004) The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2209-2214.
- Harding, S.A., Leshkevich, J., Chiang, V.L. & Tsai, C.-J. (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiology*, 128, 428-438.
- Harkess, A., Zhou, J., Xu, C. et al. (2017) The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature Communications*, 8, 1279.
- Huang, X., Hu, X., Liu, Q. et al. (2024) Full-length agave transcriptome reveals candidate glycosyltransferase genes involved in hemicellulose biosynthesis. *International Journal of Biological Macromolecules*, 274, 133508.
- Huang, X., Xiao, M., Xi, J. et al. (2019) De novo transcriptome assembly of *Agave* H11648 by Illumina sequencing and identification of cellulose synthase genes in *Agave* species. *Genes*, 10, 103.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587-589.
- Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772-780.
- Kienow, L., Schneider, K., Bartsch, M. et al. (2008) Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from *Arabidopsis thaliana*. *Journal of Experimental Botany*, 59, 403-419.
- Kuhn, D.N., Chappell, J., Boudet, A. & Hahlbrock, K. (1984) Induction of phenylalanine ammonia-lyase and 4-coumarate:CoA ligase mRNAs in cultured plant cells by UV light or fungal elicitor. *Proceedings of the National Academy of Sciences of the United States of America*, 81, 1102-1106.
- Lavhale, S.G., Kalunke, R.M. & Giri, A.P. (2018) Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta*, 248, 1063-1078.
- Le Gall, H., Philippe, F., Domon, J.-M., Gillet, F., Pelloux, J. & Rayon, C. (2015) Cell wall metabolism in response to abiotic stress. *Plants*, 4, 112-166.
- Lee, D. & Douglas, C.J. (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family: cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *Plant Physiology*, 112, 193-205.
- Li, M., Guo, L., Wang, Y. et al. (2022) Molecular and biochemical characterization of two 4-coumarate: CoA ligase genes in tea plant (*Camellia sinensis*). *Plant Molecular Biology*, 109, 579-593.
- Li, S.-S., Chang, Y., Li, B., Shao, S.-L. & Zhang, Z.-Z. (2020a) Functional analysis of 4-coumarate: CoA ligase from *Dryopteris fragrans* in transgenic tobacco enhances lignin and flavonoids. *Genetics and Molecular Biology*, 43, e20180355.
- Li, S.S., Li, Y., Sun, L.L., Hu, B.Z. & Chang, Y. (2015) Identification and expression analysis of 4-coumarate:coenzyme A ligase gene family in *Dryopteris fragrans*. *Cellular and Molecular Biology*, 61, 25-33.
- Li, W., Wang, K., Chern, M. et al. (2020b) Sclerenchyma cell thickening through enhanced lignification induced by *OsMYB30* prevents fungal penetration of rice leaves. *The New Phytologist*, 226, 1850-1863.
- Liu, L., Geng, P., Jin, X. et al. (2023) Wounding induces suberin deposition, relevant gene expressions and changes of endogenous phytohormones in chinese yam (*Dioscorea opposita*) tubers. *Functional Plant Biology*, 50, 691-700.
- Liu, X.-Y., Wang, P.-P., Wu, Y.-F., Cheng, A.-X. & Lou, H.-X. (2018) Cloning and functional characterization of two 4-coumarate:CoA ligase genes from *Selaginella moellendorffii*. *Molecules*, 23, 595.
- Livak, K.J. & Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_t} method. *Methods*, 25, 402-408.
- Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D. & Hahlbrock, K. (1988) Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate:CoA ligase genes in parsley. *European Journal of Biochemistry*, 176, 661-667.
- Ma, B., Song, W., Zhang, H. et al. (2026) Characterization of the *Nt4CL* gene family in tobacco (*Nicotiana tabacum*) and reveals *Nt4CL9* response to cold stress. *Industrial Crops and Products*, 242, 122870.
- Ma, J., Zuo, D., Zhang, X. et al. (2024) Genome-wide identification analysis of the 4-coumarate:CoA ligase (4CL) gene family expression profiles in *Juglans regia* and its wild relatives *J. mandshurica* resistance and salt stress. *BMC Plant Biology*, 24, 211.
- Ma, Z.-H., Nan, X.-T., Li, W.-F., Mao, J. & Chen, B.-H. (2023) Comprehensive genomic identification and expression analysis 4CL gene family in apple. *Gene*, 858, 147197.
- Minh, B.Q., Schmidt, H.A., Chernomor, O. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530-1534.
- Nie, T., Sun, X., Wang, S., Wang, D., Ren, Y. & Chen, Q. (2023) Genome-wide identification and expression analysis of the 4-coumarate:CoA ligase gene family in *Solanum tuberosum*. *International Journal of Molecular Sciences*, 24, 1642.
- Ran, F., Xiang, C., Wang, C. et al. (2024) Identification of the 4CL family in cassava (*Manihot esculenta* Crantz) and expression pattern analysis of the *Me4CL32* gene. *Biochemical and Biophysical Research Communications*, 735, 150731.
- Ren, W., Gao, T., Shao, S., Hou, B., Liao, X. & Ye, N. (2023) [Genome-wide identification and expression analysis of 4CL gene family in *Camellia sinensis*.] *Acta Botanica Boreali-Occidentalia Sinica*, 43, 1459-1469. [In Chinese]
- Schneider, K., Hövel, K., Witzel, K. & Stübke, H.-P. (2003) The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 8601-8606.
- Schröder, J. (1989) Protein sequence homology between plant 4-coumarate:CoA ligase and firefly luciferase. *Nucleic Acids Research*, 17, 460.
- Sha, S., Mu, K., Luo, Q. et al. (2025) Genome-wide identification and characterisation of the 4-coumarate-CoA ligase (4CL) gene family in *Gastrodia elata* and their transcriptional response to fungal infection. *International Journal of Molecular Sciences*, 26, 7610.
- Soubeyrand, E., Kelly, M., Keene, S.A. et al. (2019) *Arabidopsis* 4-COUMAROYL-COA LIGASE 8 contributes

- to the biosynthesis of the benzenoid ring of coenzyme Q in peroxisomes. *Biochemical Journal*, 476, 3521-3532.
- Sun, H., Li, Y., Feng, S. et al. (2013) Analysis of five rice 4-coumarate:coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochemical and Biophysical Research Communications*, 430, 1151-1156.
- Toljamo, A., Blande, D., Kärenlampi, S. et al. (2016) Reprogramming of strawberry (*Fragaria vesca*) root transcriptome in response to *Phytophthora cactorum*. *PLoS ONE*, 11, e161078.
- Tugbaeva, A., Ermoshin, A., Wuriyanghan, H., Maleva, M., Borisova, G. & Kiseleva, I. (2022) Copper stress enhances the lignification of axial organs in *Zinnia elegans*. *Horticulturae*, 8, 558.
- Untergasser, A., Cutcutache, I., Koressaar, T. et al. (2012) Primer3 – new capabilities and interfaces. *Nucleic Acids Research*, 40, e115.
- Wang, X.X., Huang, X, Chen, L. et al. (2023) Transcriptome sequencing of *Agave amaniensis* reveals shoot-related expression patterns of *expansin A* genes in *Agave*. *Plants*, 12, 2020.
- Wang, X.X., Huang, X.L., Chen L. et al. (2024) *Agave schidigera* transcriptome reveals stress-responsive *phenylalanine ammonia-lyase* genes in *Agave*. *Agronomy*, 14, 2520.
- Wang, Y., Fan, J., Guo, Z. et al. (2025) The *Zm4CL2* gene regulates drought stress response in *Zea mays* L. *Plant Physiology and Biochemistry*, 229, 110389.
- Wang, Y., Guo, L., Zhao, Y., Zhao, X. & Yuan, Z. (2022) Systematic analysis and expression profiles of the 4-coumarate:CoA ligase (*4CL*) gene family in pomegranate (*Punica granatum* L.). *International Journal of Molecular Sciences*, 23, 3509.
- Zhang, M., Tian M., Weng Z. et al. (2025a) Genome-wide identification analysis of the 4-coumarate:CoA ligase (*4CL*) gene family in *Brassica* U's triangle species and its potential role in the accumulation of flavonoids in *Brassica napus* L. *Plants*, 14, 714.
- Zhang, X.H. & Chiang, V.L. (1997) Molecular cloning of 4-coumarate:coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. *Plant Physiology*, 113, 65-74.
- Zhang, Z., Wang, Y. & Li, R. (2025b) Genome-wide analysis of *Zm4CL* genes identifies *Zm4CL8* regulating drought and salt tolerance in maize. *Agronomy*, 15, 1100.
- Zhang, Z., Yang, X., Ning, D. & Li, R. (2025c) Characterization of 4-coumarate-CoA ligase (*4CL*) genes in wheat uncovers Ta4CL91's role in drought and salt stress adaptation. *Plants*, 14, 1301.