Systematic Analysis and Biochemical Characterization of the Caffeoyl Shikimate Esterase Gene Family in Poplar

Nan Chao

Jiangsu University of Science and Technology

Qi Qi

China Agricultural University

Xue-Chun Wang

Beijing Forestry University

Shuang Li

Beijing Forestry University

Xiang-Ning Jiang

Beijing Forestry University

Ying Gai (✉ gaiying@bjfu.edu.cn)

College of Biological science and technology, beijing forestry university <https://orcid.org/0000-0002-1848-6040>

Research Article

Keywords: caffeoyl shikimate esterase, enzymatic assay, gene family, lignin, Populus

Posted Date: June 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-408828/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Caffeoyl shikimate esterase (CSE) hydrolyzes caffeoyl shikimate into caffeate and shikimate in the phenylpropanoid pathway. In this study, we performed systematic analysis of *CSE* gene family in poplar and investigated the possible roles of CSEs and CSE-like genes in *Populus*. We performed a genome-wide analysis of the CSE family, including functional and phylogenetic analyses of CSE and CSE-like genes using the poplar (*Populus trichocarpa*) genome. Eighteen CSE and CSE-like genes were identified in the *Populus* genome and five phylogenetic groups were identified from phylogenetic analysis. CSEs in Group Ia, which were proposed as bona fide CSEs, have probably been lost in most monocots except *Oryza sativa*. Primary functional classification showed that *PoptrCSE1* and *PoptrCSE2* had putative function in lignin biosynthesis. In addition, *PoptrCSE2*, along with *PoptrCSE12* might also respond to stress with a function in cell wall biosynthesis. Enzymatic assay of *Populus tomentosa* (Popto) CSE1, -2 and -12 showed that PoptoCSE1 and -2 kept CSE activity. PoptoCSE1 and 2 had similar biochemical properties, tissue expression pattern and subcellular localization. Most of the PoptrCSE-like genes are homologs of AtMAGL (monoacylglycerol lipase) genes in *Arabidopsis* and may function as MAG lipase in poplar. Our study provides systematic understanding of this novel gene family and suggests the CSE function in monolignol biosynthesis in *Populus*.

Introduction

Lignin is one of the most abundant biomass components in plants and has been extensively studied for decades because its vital role in the formation of the plant secondary cell wall and wood (Poovaiah et al. 2014; Vanholme et al. 2010; Bhuiyan et al. 2009; Wang et al. 2018b). The recalcitrance of lignin restricts the both forage utilization and processing of biomass for biofuels (Ragauskas et al. 2014; Zhao and Dixon 2014). Therefore to overcome this limitation, genetic modification strategies targeting on key genes involved in monolignol biosynthesis are being developed to obtain plants with less lignin or with more easily degradable lignin while maintaining normal biomass (Tang and Tang 2014; Vanholme et al. 2012; Li and Chapple 2010; Voelker et al. 2011; Stewart et al. 2009; Naik et al. 2018; Wang et al. 2018b).

Monolignol biosynthesis pathway is a branch of the phenylpropanoid pathway and has been continuously revised over the past decades. Now 11 enzyme families and 24 metabolites are involved in monolignol biosynthesis (Vanholme et al. 2019). The general phenylpropanoid pathway contains phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H) and 4-coumarate: CoA ligase (4CL) and provides hydroxycinnamoyl-CoA esters as precursors for a wide range of end products. These hydroxycinnamoyl-CoA esters were further catalyzed into monolignols by series of enzymes in specific monolignol biosynthesis pathway (Zhao 2016). Changing the mass flux in the phenylpropanoid pathway has become a new strategy for modifying lignin content (Whetten and Sederoff 1995; Boerjan et al. 2003; Vanholme et al. 2010; Weng and Chapple 2010; Vanholme et al. 2013; Ragauskas et al. 2014). Recently, dynamic flux modeling for monolignol biosynthesis was also proposed based on systematic understanding of monolignol biosynthesis pathway (Wang et al. 2018a; Guo et al. 2018; Faraji et al. 2018a; Faraji et al. 2018b). The discovery of caffeoyl shikimate esterase (CSE) as a central enzyme in the lignin biosynthetic pathway promoted a revision of our understanding of the pathway (Vanholme et al. 2013). *Arabidopsis thaliana cse* mutants have shown that lignin content decreases with increasing levels of *p*-hydroxyphenyl units and the efficiency of the conversion of cellulose to glucose in *cse* mutants is improved after saccharification without pretreatment. CSE together with 4CL directs the flux away from H lignin and toward G and S lignin. Down-regulation of CSE in hybrid poplar (*Populus tremula* × *Populus alba*) resulted in up to 25% reduced lignin deposition, increased levels of *p*-hydroxyphenyl units in the lignin polymer, and a relatively higher cellulose content (Saleme et al. 2017).

AtCSE (At1g52760) was first reported as Caffeoyl Shikimate Esterase by Vanholme et al. (2013) in *Arabidopsis* because of its ability to convert caffeoyl shikimate into caffeate (Vanholme et al. 2013). Both *in vitro* enzymatic analyses and genetic *cse* mutant confirmed its vital function in monolignol biosynthesis in *Arabidopsis*. Although CSE orthologs were firstly reported based on phylogenetic analysis in several plants such as poplar, it is confused that protein extracts from secondary differentiating xylem of *P. trichocarpa* showed no detectable CSE activity and the similar results were also reported in *Panicum virgatum* and *Oryza sativa* (Wang et al. 2014). In a recent study, detectable CSE activity was reported in crude extraction of *Panicum virgatum* and purified recombinant CSE proteins from *Medicago truncatula* and *Populus deltoides* (Ha et al. 2016). The authors also indicated that reaction catalyzed by CSE may not be essential for lignification in all plant species since they found that crude protein extracts from stems of *Brachypodium distachyon* and *Zea mays* which have no orthologs of the currently characterized CSE genes, exhibit only a weak esterase activity with caffeoyl shikimate (Ha et al. 2016). Two bona fide CSEs were characterized in hybrid poplar and down-regulation of CSE in hybrid poplar resulted in the accumulation of feruloyl shikimate and sinapoyl shikimate indicating that these compounds also are *in vivo* CSE substrates. In addition, genuine CSE has also been characterized in *Pinus* in our previous study (Wang et al. 2019).

It is necessary to consider all key genes and their homologs involved in monolignol biosynthesis when to establish dynamic flux model based on enzymatic activity since both the functional redundant and complementation homologs would affect the flux model (Wang et al. 2018a; Wang et al. 2014). Therefore, using the poplar (*Populus trichocarpa*) genome which has complete genome information and annotation we performed a genome-wide analysis of the CSE family. Based on the combining analysis of expression profile and promoter analysis, we proposed functions of these genes. Biochemical characterization of the putative bona fide CSE homologs from *Populus tomentosa* was also performed *in vitro* to confirm their catalytic function. Our study provides further understanding of this novel gene family and new target genes for modifying phenylpropanoid pathway.

Materials And Methods

Materials

Buds, cambium and roots of six-year-old *P. tomentosa* 741 were harvested as plant material in this research from Hebei, China. Samples collected were immediately frozen in liquid nitrogen and then stored in -80 °C until use. These samples from different tissues were used for s-qRT-PCR (semi-quantitative reverse transcription and polymerase Chain reaction).

Genome-wide identification of CSE gene family members

To identify the CSE sequences in *Populus*, we first performed a BlastP search in the NCBI database using AtCSE1 as a query (Coordinators 2017). Sequences from selected species with high scores, query cover >90%, identity >50%, and low E-value ($E=100$) were collected. The results of the HMMsearch were further screened based on scores and E-values ($E=5$) of both the domain and full sequence. Finally, the screening sequences were submitted to the Superfamily 1.75 database (<http://supfam.cs.bris.ac.uk/SUPERFAMILY>) to further determine their superfamily and family in the SCOP database (**Supplementary Table S1**). The CSE sequences used for phylogenetic analysis (**Supplementary Table S2**) were screened from various sources based on the platform PLAZA 3.0 (<http://bioinformatics.psb.ugent.be/plaza/>). CSE sequences from the gymnosperm *Picea sitchensis* and the fern *Selaginella moellendorffii* were obtained using BlastP in the NCBI database. The rules that were proposed by Raes et al. (Raes et al. 2003) to distinguish bone fide genes and “likes” were used to determine bone fide CSE and CSE-like genes.

Organization and distribution of CSE genes and homologs on *Populus* chromosomes

A wheel diagram of the chromosome-level reorganization of the most recent genome-wide duplication event in *Populus* was created. Based on the *Populus* genome, CSE and CSE homologs were located on their respective chromosomes in specific duplicated blocks. Duplicated blocks were determined using WGDplot in the PLAZA platform. A schematic diagram was also constructed to describe the organization and distribution of CSE and its homologs on *Populus* chromosomes.

Promoter analysis of PoptrCSE (*P. trichocarpa*) genes

Promoter sequences of PoptrCSE genes were extracted using PLAZA platform with parameter upstream 1.5kb. Promoter analysis was performed by querying all CSE genes for 1500 bp upstream or up to the closest gene using the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) with selected matrix score ³5 for all elements (Lescot et al. 2002). SNBE and SMRE were detected based on the motif sequences reported by Zhong et al. (Zhong and Ye 2012). Details are presented in **Supplementary Table S3**.

CSE sequence alignment and phylogenetic analysis

Both DNAMAN 8.0 (Lynnon BioSoft) and CLUSTAL W software assembled in Mega 5.05 software were used to align candidate PoptrCSEs and bona fide CSEs using default parameters (Tamura et al. 2011). A phylogenetic tree was obtained using Mega 5.05 with the maximum-likelihood method and the JTT substitution model and G+I rates among sites model. The reliability of internal branches was assessed by bootstrapping using 1000 bootstrap replicates and marked above nodes only if greater than 50. The putative CSE protein sequences used are listed in **Supplementary Table S2**.

RNA extraction and cDNA synthesis

Total RNA from different tissues was extracted using Plant RN38 Kit (Aidlab, Beijing, China) according to the manual. The quality of RNA was evaluated by both Nanodrop one (Thermo Scientific, Waltham, Massachusetts, USA) and electrophoresis. cDNA was synthesized with the PC54-TRUEscript RT kit (Aidlab, Beijing, China) according to the manufacturer's protocol.

CSE expression profile in *P. trichocarpa* and *P. tomentosa*

Gene expression profile for various tissues including developing phloem (DP), developing xylem (DX), mature leaf (ML), and shoot and leaf primordium (SLp) in *P. trichocarpa* was performed using the GEO database with the accession number GSE30507. Transcriptome datasets of these samples were re-analyzed by submitting to Pop's Pipes: Poplar Gene Expression Pipelines (<http://sys.bio.mtu.edu/>). In addition, RNA-seq dataset GSE78953 was also used for confirming CSE gene expression and co-expression based on person correlation with genes involved in lignin biosynthesis was also performed for exploring the putative function of PoptrCSE and PoptrCSE-like genes. GSE78953 contains the transcriptome information of xylem tissue of transgenic and wildtype *P. trichocarpa*. MultiExperiment Viewer (MeV) version 4.9 and Cytoscape 3.4 were used to visualize the CSE expression profile or co-expression network. The detailed information is available in **Supplementary Table S5**. Semi-qRT-PCR was performed for *PoptoCSE1*, *PoptoCSE2* and several other monolignol biosynthesis genes in *P. tomentosa*. Actin was used as reference gene.

Subcellular location of PoptoCSE

The method used for subcellular location of PoptoCSE is according to our previous study (Wang et al. 2019). The pBI121 vector with *PoptoCSE1* and *PoptoCSE2* gene and green fluorescent protein (GFP) fusion expression was constructed by restriction enzymes and then ligation with T4 ligase (Takara, Dalian), and then the recombinant plasmids pBI121-35S-*PoptoCSE1* and pBI121-35S-*PoptoCSE2* were transferred into *Agrobacterium tumefaciens* strain GV3101 which were transferred into tobacco leaves via *Agrobacterium*-mediated transient transformation (Morita et al. 2015). The tobacco leaves were cut into squares of 5 mm × 5 mm and GFP fluorescence in leaves was observed using a Leica TCS SP8 confocal microscope (Leica Microsystems, Wetzlar, Germany).

Cloning and purification of recombinant CSEs and HCT (*p*-hydroxycinnamoyl-CoA: quinate/shikimate *p*-hydroxycinnamoyl transferase) from *P. tomentosa*

To obtain caffeoyl shikimate, we cloned both homologous CSEs and HCT from *P. tomentosa* based on the sequence information from *P. trichocarpa*. HCT can catalyze caffeoyl-CoA and shikimic acid to produce caffeoyl shikimate which is the substrate of CSEs. Primer pairs were listed in Supplementary Table S7. The candidate CSE genes identifying as genuine CSEs involved in lignin biosynthesis and HCT6 were constructed with pET28a (Novagen) and then transformed into *E. coli* BL21(DE3). The recombinant proteins tagged with His were purified using Ni-NTA agarose (Qiagen). After washing the column using lysis buffer with gradient concentrations of imidazole, the His-tagged PoptoCSEs and PoptoHCT were eluted with 100mM imidazole in lysis buffer finally. The

eluted proteins were detected by SDS-PAGE and stained by Coomassie brilliant blue R250. Concentration of purified protein was determined using Bradford Protein Assay Kit (Sangon Biotech, Shanghai).

Enzymatic assay of recombinant PoptoCSEs

We obtained caffeoyl shikimate using enzymatic reaction catalyzed by purified recombinant PoptoHCT6 with caffeoyl-CoA and shikimic acid. The reaction was performed according to the methods providing by Escamilla-Treviño *et al.* (2014). Caffeoyl-CoA was chemically synthesized and authorized as reported in our previous study (Chao *et al.* 2016). Purified PoptoCSEs were incubated at 30 °C for 30 min with 100 mM NaPO₄ buffer pH 7.5, 500 µM dithiothreitol and 100 µM caffeoyl shikimate in a final volume of 100 µL (Vanholme *et al.* 2013; Saleme *et al.* 2017; Ha *et al.* 2016). The controls were also performed with proteins that have been boiled for 10 min. The reactions (including controls) were terminated by adding 10 µL of glacial acetic acid. CSE activity was monitored by the appearance of caffeic acid. To explore the optimum reaction temperature, the reactions were incubated at various temperatures ranged from 10 °C to 50 °C. Multi-pH phosphate buffer ranged from 4 to 9 were adopted for exploring the optimum reaction pH for CSE. Reaction products with 100 ng sinapic acids which were added as internal standard were further purified using microporous filters and then injected onto an HPLC with a reverse-phase C18 column (ZORBAX 300SB-C18, 2.1 × 150 mm, 3.5 µm; Agilent, Santa Clara, CA, USA) and separated in a step gradient using 1% formic acid in water as solvent A and acetonitrile as solvent B. The gradient profile was 100% A, decreased to 65 % A in 24 min, then to 0 % A in 3 min and maintained for 5 min, and increased to 100 % A in 3 min and maintained for 20 min. The flow rate was 0.15 mL/min. The substrates and products were identified using an ion trap mass spectrometer (LCQ DECA XP MAX) coupled with an ESI source (Thermo Finnigan). The parameters used for MS analysis was sheath gas (nitrogen) flow rate, 40 arb; aux/sweep gas (nitrogen) flow rate, 10 arb; spray voltage, 4.5 keV; capillary temperature, 320 °C.

GenBank accession number

The Genbank accession numbers of PtoHCT6 and PtoCSE1, -2 and -12 are KT020989, KT020990, KT020997 and KT021003, respectively.

Statistical analyses

The correlation analysis and ANOVA analysis were performed using SPSS19.0. Significance ($p < 0.05$) was marked.

Results

Organization and distribution of the CSE gene family

We found that the number of CSE gene family members differed among the examined species and the CSE gene family may have experienced duplication or loss. Eighteen CSE homologs were identified in poplar. These PoptrCSE or PoptrCSE-like genes are distributed on 13 chromosomes (**Fig. 1**), indicating that block duplication resulting from 'Salicoid' genome duplication may have played a role in the formation of the CSE family (Tuskan *et al.* 2006). Four chromosomes, including chromosomes I, VIII, X, and XI, contain two or three PoptrCSE genes, most of which are located in homologous duplicated blocks and form 10 duplicated pairs. *PoptrCSE3* is located on a duplicated segment on chromosome XI without a matching duplicate. It is possible that its duplicate was lost during evolution. Ks analysis of the duplicated CSE gene pairs showed that 7 of the 10 CSE gene pairs experienced a recent duplication event with Ks values between 0.2 and 0.4 (**Supplementary Table S6**).

The CSE gene family was also screened from a number of plant species, including mosses (*P. patens*), ferns (*S. moellendorffii*), gymnosperms (*P. sitchensis*), and angiosperms in present study. Eighteen PoptrCSE and PoptrCSE-like genes were identified in *P. trichocarpa*. These PoptrCSE genes were distributed on 13 chromosomes with homologous pairs, except *PoptrCSE3* which was also located in a duplicate block. Inter-species collinear analysis revealed that *PoptrCSE3* is orthologous to the *AtrCSE1* gene in *Amborella* while we can't find its positional paralog relative to the duplicated blocks in *Populus* (**Fig. 2b**). In addition, based on the aligned gene strings of a set of homologous segments with *PoptrCSE11* and *PoptrCSE13* and the fact that *PoptrCSE13* has only a 361-bp upstream sequence, we propose that it is an insertion of a new gene (PT10G22220) into the promoter region of *PoptrCSE13* that leads to this change in the *PoptrCSE13* promoter region (**Fig. 2a**). The triangular connection among *PoptrCSE10*, -12, and -18 with high Ks value should be the results of the ancient 'Salicoid' genome duplication and following fission and fusion of chromosomes (Dai *et al.* 2014; Hou *et al.* 2016; Tuskan *et al.* 2006). We can still observe the collinear relationships among *AtrCSE5*-*PoptrCSE12*-*PoptrCSE18* as well as the loss of genes in gene strings on the homologous segments (**Fig. 2c**). The duplication block containing *PoptrCSE10* is homologous to that with *PoptrCSE18*, indicating that the duplication event occurred between *PoptrCSE10* and *PoptrCSE18*.

Promoter sequence analysis

All of the upstream 1500-bp sequences of PoptrCSE and PoptrCSE-like genes were obtained, except for *PoptrCSE13*, which has only 361 bp to its closest gene. Responsive elements for the six types of hormones were found in different PoptrCSE genes (**Table 1**) (Song and Wang 2009; Ogawa *et al.* 2003; Hobo *et al.* 1999; Hao *et al.* 1998; Shah and Klessig 1996; Guilfoyle *et al.* 1993). These hormones are important for plant development and defense. AC-rich elements corresponding to the MYB transcription factor-binding motif are necessary for vascular expression (Hartmann *et al.* 2005; Hatton *et al.* 1995). PoptrCSE1, with the highest similarity to AtCSE1, had both the AC-I and AC-II elements in its promoter region. *PoptrCSE2* and *PoptrCSE12* had the AC-I element, and PoptrCSE5 and 6 had the AC-II element. In addition, motifs binding to different MYBs were found in many PoptrCSE genes. Among these PoptrCSE genes, PoptrCSE18 had two MYB-binding motifs, MBS and MBSI, involved in drought response and flavonoid biosynthetic gene regulation, respectively. Some members of the MYB family have been reported as lignin-specific transcription factors (Zhao and Dixon 2011; Legay *et al.* 2010; Öhman *et al.* 2013; Fornale *et al.* 2006). In addition, we identified a secondary wall NAC binding element (SNBE) and a secondary wall MYB responsive element (SMRE) in the promoter regions of PoptrCSE and PoptrCSE-like genes, except *PoptrCSE13* (**Supplementary Table S3**). Both SNBE and SMRE were reported to be involved in secondary cell wall

biosynthesis (McCarthy et al. 2011; Zhong and Ye 2014). Furthermore, *cis*-elements involved in the stress response were revealed and were widely detected in *PoptrCSE2* and *PoptrCSE12* promoter regions.

Alignment and phylogenetic analysis of CSE and CSE-like genes

Detailed analyses of putative protein sequences of *PoptrCSE* and *AtCSE1* were also performed (Fig.3). The acyltransferase (HX₄D) and two hydrolase (GXSG) motifs were identified in these putative proteins (Vijayaraj et al. 2012; Gao et al. 2010). Two GXSG motifs were highly conserved in all *PoptrCSE* proteins. HX₄D, which was reported as a key motif for acyltransferase activity for monoacylglycerol acyltransferase (MGAT) in peanuts, is less conserved (Vijayaraj et al. 2012). While aspartate and glutamate residues in the sixth position are present in most *PoptrCSE* proteins, this position was substituted by threonine in *PoptrCSE10*, -12, and -18. *PoptrCSE1*, -2, and -3 kept D in this position as same as *AtCSE1*. The catalytic triad (S, D, and H), which was reported to be responsible for lipase/esterase activity, was identified as highly conserved in these proteins (black circles) (Wang et al. 1997). Another two proteins, *PoptrCSE14* and *PoptrCSE18*, were motif-deficient. *PoptrCSE14* only had the first hydrolase motif without the other two motifs and *PoptrCSE18* lacked the first hydrolase motif and even appeared to have an incomplete catalytic triad because of its shorter peptide sequence. Above variations indicate a functional change or deficiency in these two proteins.

Phylogenetic analysis was performed on a total of 153 putative CSE and CSE-like proteins including *AtMAGLs* (monoacylglycerol lipase) from the 18 species. The hydrolase (GXSG) motifs in these putative proteins were highly conserved, corresponding to the lipase/esterase superfamily, whereas the HX₄D motifs experienced divergence, especially in the D position, which can be used to reveal the evolutionary relationships between these CSEs. Based on the evolution of the acyltransferase motif and the maximum likelihood (ML) bootstrap phylogenetic tree (Fig. 4), all of these putative CSE or CSE-like proteins were divided into five groups. It is noteworthy that most of the CSE or CSE-like proteins in Group I have the classical acyltransferase motif with HX₄D. Further analysis indicated that this group could be divided into two subsets. Group Ia, is herein classified as a bona fide CSE group based on the rules described by Raes *et al.* (2003). Putative proteins in this subgroup have high sequence identity with *AtCSE1*, which was identified as a bona fide CSE in the functional analyses (Vanholme et al. 2013). *PoptrCSE1* and 2, bona fide *PoptrCSEs*, were grouped with *AtCSE1* in Group Ia and are likely involved in lignin biosynthesis with CSE activity. It is interesting that bona fide CSEs, except two genes from *Oryza sativa* (*OsCSE5* and *OSINDICA_CSE7*), are found primarily in dicots. CSEs in Group Ib are relatively more closely related to bona fide CSEs. Almost all of the putative proteins in Group II and Group III have HX₄E instead of HX₄D. Group IV refers to CSEs with HX₄T/S. The last group (Group V) contains *PoptrCSE15*, -16, -17 and -9 and mainly has HX₄H instead of the acyltransferase motif. The catalytic residues of putative CSEs are in **Supplementary Table S4**, grouped according to the phylogenetic tree.

Expression analysis of CSE genes

From the results of the hierarchical cluster analysis based on expression levels in different tissues or organs, three expression clusters were found (Fig. 5a). *PoptrCSE4*, -7, and -10 show obvious expression preferences in mature leaves and clustered with *PoptrCSE 5* and -13, in cluster I with relatively higher expression abundance in mature leaves (ML). *PoptrCSE* and *PoptrCSE*-like genes in cluster II had relatively high expression levels in shoots and leaf primordium (SLp) with no obvious preferences in all detected tissues and organs. *PoptrCSE1*, -2, -9, -12, and -17, constituting cluster III, showed expression preferences in developing phloem and xylem, which are involved in wood formation. Co-expression network shows that *PoptrCSE1*, -2 have significant correlation with genes involved in lignin biosynthesis, which can be used as evidence to explore the putative function of these *PoptrCSEs* (Fig. 5b).

Characteristic of PoptoCSE1 and PoptoCSE2

We further cloned and obtained the corresponding *PoptoCSE1*, -2 and -12 from *P. tomentosa*. The *PoptoCSE1*, -2 and -12 have high sequence identity (>98%) with corresponding CSEs in *P. trichocarpa*. sqRT-PCR showed that *PoptoCSE1*, *PoptoCSE2* have similar expression pattern to other monolignol biosynthesis related genes such as *PoptoCAD*, *PoptoCCR7*, *PoptoCCoAOMT* (Fig. 5c). This fact further implies *PoptoCSE1* and *PoptoCSE2* are likely to involving in monolignol biosynthesis. Both *PoptoCSE1* and *PoptoCSE2* with fused GFP were observed in plasma membrane and ER (Fig. 6). This result is corresponding with the subcellular localization of *AtCSE* (At1g52760) which was reported as lysophospholipase 2 (lysoPL2) (Gao et al. 2010).

Enzymatic assay of PoptoCSE1, -2 and -12 of P. tomentosa

The recombinant proteins *PoptoCSE1* and *PoptoCSE2* were obtained and purified for enzymatic assay. We detected the CSE activity using HPLC-MS and found that *PoptoCSE1* and *PoptoCSE2* have detectable activity against caffeoyl shikimate (Fig.7 a, b, c) while *PoptoCSE12* showed no activity against caffeoyl shikimate (Fig. S1). In addition, we also explored the optimum reaction temperature and pH for *PoptoCSE1* and *PoptoCSE2* (Fig.7d, E). *PoptoCSE1* and *PoptoCSE2* have similar temperature profile and pH profile (except at pH 8.0-9.0) for catalyzing caffeoyl shikimate and the highest activities for *PoptoCSEs* were obtained at pH =7.0 and temperature at 40 °C. *PoptoCSE1* and *PoptoCSE2* showed similar biochemical properties against caffeoyl shikimate.

Discussion

Phylogenetic analysis and comparison of the acyltransferase motif of 153 putative CSEs suggested the existence of five CSE and CSE-like groups. *PoptrCSE1* and *PoptrCSE2* were identified as bona fide CSEs as class I. In previous study, two bona fide CSEs in *Populus* were also indicated by phylogenetic trees. (Kim et al. 2016; Gao et al. 2010) In addition, subcellular localization and the enzymatic assay of *PoptoCSE1* and *PoptoCSE2* provided further evidence to confirm the existence of genuine CSE involved in monolignol biosynthesis in poplar. CSE in genuine clade was also identified in gymnosperm which is corresponding to our previous study in *Larix Kaempferi* (Wang et al. 2019). The deficiency of CSE in most monocots are corresponding with the studies in *Brachypodium distachyon* and *Zea may* (Ha et al. 2016). It is possible that ancestral CSE appears before the divergence of angiosperms and gymnosperms. The bona fide CSEs (Group Ia) were likely lost in most monocots at the divergence between dicots and monocot. *AtCSE* was first annotated as a lysoPL2 enzyme and also

identified as member of monoacylglycerol lipase (MAGL) gene family (Kim et al. 2016; Gao et al. 2010). Here we further classified these CSE-like proteins into four groups with different amino acid replacement at the last position of acyltransferase motif. Phylogenetic analysis also showed these PoptrCSE-like proteins are clustered with AtMAGLs (Fig. 4 and Fig. S2).

Based on above analysis in present study, we primarily proposed the functions of PoptrCSE and PoptrCSE-like genes. *PoptrCSE1**PoptrCSE2* and *-12* are expressed mainly in developing xylem (cluster III). Co-expression analysis shows that *PoptrCSE1*, *-2* and *-12* have significant correlation with genes involved in lignin biosynthesis. AC-I, AC-II, SMREs and SNBEs are involved in regulating secondary cell wall biosynthesis (Zhong and Ye 2014; Zhao and Dixon 2011). It has been reported that MYB58 or MYB63 can bind AC element, the conserved promoter sequence in most monolignol biosynthetic genes to directly activate the expression of these genes in *A. thaliana* (Zhou et al. 2009). *PoptrCSE1*, *-2* and *-12* all contain the AC element (AC-I or AC-II) in the promoter regions. SMREs are additional cis elements in addition to the AC elements and Arabidopsis MYB46/83, master switches of the entire secondary wall biosynthetic program are able to bind to SMREs (Zhong and Ye 2012). SECONDARY WALL NACs (SWNs), master transcriptional switches activating secondary wall biosynthesis were reported to bind to SNREs and their binding to the SNBE sites were proposed to be necessary for a combinatorial activation of the secondary wall biosynthesis genes with other transcription factors (Zhong et al. 2010). In our promoter analysis, SMREs and SNREs were found to widely distribute in most CSE and CSE-like gene promoter regions and *PoptrCSE1* and *-2* contain the most SMREs and SNBEs (Table1). Although *PoptrCSE1* and *-2* have high sequence identity with *AtCSE1* and both have many cis-elements involved in secondary wall biosynthesis, the promoter analysis revealed that *PoptrCSE2* and *-12* have many cis-elements involved in stress response that differ from *PoptrCSE1*. The above factors indicate that the *PoptrCSE1*, *-2* and *-12* are likely involved in wood formation, but *PoptrCSE2* and *PoptrCSE12* might respond mainly to stress. This functional divergence has been reported in other key genes involved in lignin biosynthesis (Tronchet et al. 2010; Moura et al. 2010; Lauvergeat et al. 2001). PoptrCSE3 belonging to phylogenetic Group Ib showed close relationship with AtGAML1 which was reported to harbor MAG lipase activities and lysophosphatidylcholine (LPC) and/or lysophosphatidylethanolamine (LPE) hydrolase activities. PoptrCSE3 and AtGAML1 have both acyltransferase (HX4D) and hydrolase (GXSXG) motifs and are proposed to have acyltransferase activity. PoptrCSE6, *-7*, *-9*, *-14*, *-15*, *-16*, and *-17* from phylogenetic group III and V all have cis-elements involved in response to hormones. Previous studies have revealed the bifunction of AtCSE and its function in phospholipid metabolism (Vijayaraj et al. 2012; Gao et al. 2010). PoptrCSE9, *-15*, *-16*, and *-17* cluster with AtMAGL2 and *-4* which were reported to have MAG lipase activities and LPC and/or LPE hydrolase activities. In addition, expression of AtMAGL2 and *-4* was significantly induced by salt, osmotic and cold stresses. Roles of lipases in jasmonate metabolism and phosphatidic acid signaling associated with plant defense have also been reported (Shah 2014). Therefore, CSE-like proteins in this class likely function as lipases and are here thought to be involved in plant defense in response to salicylic acid or MeJA. Other PoptrCSE-like genes also have corresponding homologous AtMAGLs and might also have MAG lipase activity.

Declarations

Funding

This work was jointly supported by the Beijing Higher Education Young Elite Teacher Project [YETP0755 granted to Dr. YING GAI], the National Natural Science Foundation [NSF 31300498 to Y.G.], and the Fundamental Research Funds for the Central Universities [BLYJ201504 to N.C.].

Author Contributions

Ying Gai and Xiang-Ning Jiang guided the work and provided advices; Qi Qi, Xu-Chun Wang, Shuang Li and Nan Chao performed the experiments; Nan Chao analyzed the data and wrote the paper.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Bhuiyan NH, Selvaraj G, Wei Y, King J (2009) Role of lignification in plant defense. *Plant Signal Behav* 4 (2):158-159
- Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annu Rev Plant Biol* 54:519-546. doi:10.1146/annurev.arplant.54.031902.134938
- Chao N, Li N, Qi Q, Li S, Lv T, Jiang XN, Gai Y (2016) Characterization of the cinnamoyl-CoA reductase (CCR) gene family in *Populus tomentosa* reveals the enzymatic active sites and evolution of CCR. *Planta* 245:1-15
- Coordinators NR (2017) Database resources of the national center for biotechnology information. *Nucleic acids research* 45 (Database issue):D12
- Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA, Milne R, Chen Y, Wan Z, Wang Z, Luo W, Wang K, Wan D, Wang M, Wang J, Liu J, Yin T (2014) The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res* 24 (10):1274-1277. doi:10.1038/cr.2014.83
- Faraji M, Fonseca LL, Escamillatrevino L, Barrosrios J, Engle NL, Yang ZK, Tschaplinski TJ, Dixon RA, Voit EO (2018a) A dynamic model of lignin biosynthesis in *Brachypodium distachyon*. *Biotechnology for Biofuels* 11 (1):253
- Faraji M, Fonseca LL, Escamillatrevino L, Barrosrios J, Engle NL, Yang ZK, Tschaplinski TJ, Dixon RA, Voit EO (2018b) Mathematical models of lignin biosynthesis. *Biotechnology for Biofuels* 11 (1):34-34

- Fornale S, Sonbol FM, Maes T, Capellades M, Puigdomenech P, Rigau J, Caparros-Ruiz D (2006) Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors. *Plant Mol Biol* 62 (6):809-823. doi:10.1007/s11103-006-9058-2
- Gao W, Li HY, Xiao S, Chye ML (2010) Acyl-CoA-binding protein 2 binds lysophospholipase 2 and lysoPC to promote tolerance to cadmium-induced oxidative stress in transgenic *Arabidopsis*. *The Plant Journal* 62 (6):989-1003
- Guilfoyle T, Hagen G, Li Y, Ulmasov T, Liu Z, Strabala T, Gee M (1993) Auxin-regulated transcription. *Functional Plant Biology* 20 (5):489-502
- Guo L, Wang P, Jaini R, Dudareva N, Chapple C, Morgan JA (2018) Dynamic modeling of subcellular phenylpropanoid metabolism in *Arabidopsis* lignifying cells. *Metabolic Engineering* 49:36-46
- Ha CM, Escamilla-Trevino L, Yarce JC, Kim H, Ralph J, Chen F, Dixon RA (2016) An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant J* 86 (5):363-375. doi:10.1111/tpj.13177
- Hao D, Ohme-Takagi M, Sarai A (1998) Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *Journal of Biological Chemistry* 273 (41):26857-26861
- Hartmann U, Sagasser M, Mehrrens F, Stracke R, Weisshaar B (2005) Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. *Plant molecular biology* 57 (2):155-171
- Hatton D, Sablowski R, Yung MH, Smith C, Schuch W, Bevan M (1995) Two classes of cis sequences contribute to tissue-specific expression of a PAL2 promoter in transgenic tobacco. *The Plant Journal* 7 (6):859-876
- Hobo T, Asada M, Kowiyama Y, Hattori T (1999) ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. *The Plant Journal* 19 (6):679-689
- Hou J, Ye N, Dong Z, Lu M, Li L, Yin T (2016) Major Chromosomal Rearrangements Distinguish Willow and Poplar After the Ancestral "Salicoid" Genome Duplication. *Genome Biol Evol* 8 (6):1868-1875. doi:10.1093/gbe/evw127
- Kim RJ, Kim HJ, Shim D, Suh MC (2016) Molecular and biochemical characterizations of the monoacylglycerol lipase gene family of *Arabidopsis thaliana*. *Plant J* 85 (6):758-771. doi:10.1111/tpj.13146
- Lauvergeat V, Lacomme C, Lacombe E, Lasserre E, Roby D, Grima-Pettenati J (2001) Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. *Phytochemistry* 57 (7):1187-1195
- Legay S, Sivadon P, Blervacq AS, Pavy N, Baghdady A, Tremblay L, Levasseur C, Ladouce N, Lapierre C, Seguin A, Hawkins S, Mackay J, Grima-Pettenati J (2010) EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. *New Phytol* 188 (3):774-786. doi:10.1111/j.1469-8137.2010.03432.x
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic acids research* 30 (1):325-327
- Li X, Chapple C (2010) Understanding lignification: challenges beyond monolignol biosynthesis. *Plant Physiol* 154 (2):449-452. doi:10.1104/pp.110.162842
- McCarthy RL, Zhong R, Ye Z-H (2011) Secondary wall NAC binding element (SNBE), a key cis-acting element required for target gene activation by secondary wall NAC master switches. *Plant signaling & behavior* 6 (9):1282-1285
- Morita S, Yamashita Y, Fujiki M, Todaka R, Nishikawa Y, Hosoki A, Yabe C, Nakamura Ji, Kawamura K, Suwastika IN, Sato MH, Masumura T, Ogihara Y, Tanaka K, Satoh S (2015) Expression of a rice glutaredoxin in aleurone layers of developing and mature seeds: subcellular localization and possible functions in antioxidant defense. *Planta* 242 (5):1195-1206. doi:10.1007/s00425-015-2354-9
- Moura JCMS, Bonine CAV, De Oliveira Fernandes Viana J, Dornelas MC, Mazzafera P (2010) Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of Integrative Plant Biology* 52 (4):360-376
- Naik P, Wang JP, Sederoff R, Chiang V, Williams C, Ducoste JJ (2018) Assessing the impact of the 4CL enzyme complex on the robustness of monolignol biosynthesis using metabolic pathway analysis. *PLoS One* 13 (3):e0193896. doi:10.1371/journal.pone.0193896
- Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S (2003) Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *The Plant Cell Online* 15 (7):1591-1604
- Öhman D, Demedts B, Kumar M, Gerber L, Gorzsás A, Goeminne G, Hedenström M, Ellis B, Boerjan W, Sundberg B (2013) MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in *Arabidopsis* stems. *The Plant Journal* 73 (1):63-76
- Poovaiah CR, Nageswara-Rao M, Soneji JR, Baxter HL, Stewart CN (2014) Altered lignin biosynthesis using biotechnology to improve lignocellulosic biofuel feedstocks. *Plant Biotechnol J*

- Raes J, Rohde A, Christensen JH, De Peer YV, Boerjan W (2003) Genome-Wide Characterization of the Lignification Toolbox in Arabidopsis. *Plant Physiol* 133 (3):1051-1071
- Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M (2014) Lignin Valorization: Improving Lignin Processing in the Biorefinery. *Science* 344 (6185):1246843
- Saleme MLS, Cesarino I, Vargas L, Kim H, Vanholme R, Goeminne G, Van Acker R, Fonseca FCA, Pallidis A, Voorend W, Junior JN, Padmakshan D, Van Doorselaere J, Ralph J, Boerjan W (2017) Silencing CAFFEYOYL SHIKIMATE ESTERASE Affects Lignification and Improves Saccharification in Poplar. *Plant Physiol* 175 (3):1040-1057. doi:10.1104/pp.17.00920
- Shah J (2014) Lipases in Signaling Plant Defense Responses. In: *Phospholipases in Plant Signaling*. Springer, pp 207-228
- Shah J, Klessig DF (1996) Identification of a salicylic acid-responsive element in the promoter of the tobacco pathogenesis-related β -1, 3-glucanase gene, PR-2d. *The Plant Journal* 10 (6):1089-1101
- Song J, Wang Z (2009) Molecular cloning, expression and characterization of a phenylalanine ammonia-lyase gene (SmpAL1) from *Salvia miltiorrhiza*. *Molecular biology reports* 36 (5):939-952
- Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiol* 150 (2):621-635. doi:10.1104/pp.109.137059
pp.109.137059 [pii]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 28 (10):2731-2739
- Tang W, Tang AY (2014) Transgenic woody plants for biofuel. *Journal of Forestry Research* 25 (2):225-236
- Tronchet M, Balague C, Kroj T, Jouanin L, Roby D (2010) Cinnamyl alcohol dehydrogenases-C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in Arabidopsis. *Mol Plant Pathol* 11 (1):83-92
- Tuskan G, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A (2006) The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313 (5793):1596-1604
- Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science* 341 (6150):1103-1106
- Vanholme R, De Meester B, Ralph J, Boerjan W (2019) Lignin biosynthesis and its integration into metabolism. *Curr Opin Biotechnol* 56:230-239. doi:10.1016/j.copbio.2019.02.018
- Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. *Plant Physiol* 153 (3):895-905. doi:10.1104/pp.110.155119
- Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W (2012) Metabolic engineering of novel lignin in biomass crops. *New Phytol* 196 (4):978-1000
- Vijayaraj P, Jashal CB, Vijayakumar A, Rani SH, Venkata Rao DK, Rajasekharan R (2012) A Bifunctional Enzyme That Has Both Monoacylglycerol Acyltransferase and Acyl Hydrolase Activities. *Plant Physiol* 160 (2):667-683. doi:10.1104/pp.112.202135
- Voelker SL, Lachenbruch B, Meinzer FC, Strauss SH (2011) Reduced wood stiffness and strength, and altered stem form, in young antisense 4CL transgenic poplars with reduced lignin contents. *New Phytol* 189 (4):1096-1109. doi:10.1111/j.1469-8137.2010.03572.x
- Wang A, Loo R, Chen Z, Dennis EA (1997) Regiospecificity and catalytic triad of lysophospholipase I. *J Biol Chem* 272 (35):22030-22036
- Wang JP, Matthews ML, Naik PP, Williams CM, Ducoste JJ, Sederoff RR, Chiang VL (2018a) Flux modeling for monolignol biosynthesis. *Curr Opin Biotechnol* 56:187-192. doi:10.1016/j.copbio.2018.12.003
- Wang JP, Matthews ML, Williams C, Shi R, Yang C, Tunlayaanukit S, Chen H, Li Q, Liu J, Lin C (2018b) Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. *Nature Communications* 9 (1):1579
- Wang JP, Naik PP, Chen H-C, Shi R, Lin C-Y, Liu J, Shuford CM, Li Q, Sun Y-H, Tunlaya-Anukit S (2014) Complete Proteomic-Based Enzyme Reaction and Inhibition Kinetics Reveal How Monolignol Biosynthetic Enzyme Families Affect Metabolic Flux and Lignin in *Populus trichocarpa*. *The Plant Cell Online* 26 (3):894-914
- Wang X, Chao N, Zhang M, Jiang X, Gai Y (2019) Functional Characteristics of Caffeoyl Shikimate Esterase in *Larix Kaempferi* and Monolignol Biosynthesis in Gymnosperms. *Int J Mol Sci* 20 (23). doi:10.3390/ijms20236071
- Weng JK, Chapple C (2010) The origin and evolution of lignin biosynthesis. *New Phytol* 187 (2):273-285. doi:10.1111/j.1469-8137.2010.03327.x

Whetten R, Sederoff R (1995) Lignin Biosynthesis. *Plant Cell* 7 (7):1001-1013. doi:10.1105/tpc.7.7.1001

Zhao Q (2016) Lignification: flexibility, biosynthesis and regulation. *Trends Plant Sci* 21 (8):713-721

Zhao Q, Dixon RA (2011) Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci* 16 (4):227-233. doi:10.1016/j.tplants.2010.12.005

S1360-1385(10)00258-X [pii]

Zhao Q, Dixon RA (2014) Altering the Cell Wall and Its Impact on Plant Disease: From Forage to Bioenergy. *Annu Rev Phytopathol* (0)

Zhong R, Lee C, Ye ZH (2010) Global Analysis of Direct Targets of Secondary Wall NAC Master Switches in Arabidopsis. *Molecular Plant*

Zhong R, Ye Z-H (2012) MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. *Plant and Cell Physiology* 53 (2):368-380

Zhong R, Ye Z-H (2014) Complexity of the transcriptional network controlling secondary wall biosynthesis. *Plant Science* 229:193-207

Zhou J, Lee C, Zhong R, Ye Z-H (2009) MYB58 and MYB63 Are Transcriptional Activators of the Lignin Biosynthetic Pathway during Secondary Cell Wall Formation in Arabidopsis. *Plant Cell* 21 (1):248-266

Tables

Table 1 List of motifs found in the promoter regions of CSE genes in *Populus*. PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) with selected matrix score ³ 5 for all elements was adopted to obtain the putative motifs. The number of specific motif in each gene was marked. More detailed information for all detected motifs is available in **Supplementary Table S3**. SNBE: secondary wall NAC binding element, SMRE: secondary wall MYB responsive element, LTR: low temperature responsive element, HSE : heat stress responsive element.

	Number of SNBEs	Number of SMREs	Salicylic acid	MeJA	Gibberellin	Auxin	Abscisic acid	Ethylene	AC-I element	AC-II element	LTR	HSE	Anaerobic induction
PoptrCSE1	4	5	-	-	1	-	1	-	2	1	1	-	2
PoptrCSE2	4	6	-	-	-	-	1	-	1	-	3	3	-
PoptrCSE3	1	4	1	-	1	1	-	1	-	-	2	2	5
PoptrCSE4	4	1	-	-	-	-	1	-	-	-	-	-	2
PoptrCSE5	6	2	-	-	1	1	-	-	-	1	-	-	2
PoptrCSE6	4	2	2	-	1	1	1	1	-	1	-	2	3
PoptrCSE7	3	5	2	1	-	-	2	-	-	-	-	4	1
PoptrCSE8	5	3	-	1	1	-	1	2	-	-	-	-	2
PoptrCSE9	3	3	1	-	1	1	-	1	-	-	2	2	5
PoptrCSE10	2	5	1	2	-	1	1	1	-	-	-	-	3
PoptrCSE11	2	3	-	-	-	-	1	-	-	-	-	-	5
PoptrCSE12	5	2	1	1	-	-	1	1	1	-	-	3	1
PoptrCSE13	0	0	-	-	-	-	1	-	-	-	1	1	-
PoptrCSE14	7	2	1	2	1	1	1	-	-	-	-	3	1
PoptrCSE15	5	4	1	3	-	2	1	-	-	-	1	1	-
PoptrCSE16	5	4	1	3	-	-	-	-	-	-	-	2	1
PoptrCSE17	2	2	1	2	2	1	3	1	-	-	1	1	-
PoptrCSE18	2	4	1	1	1	-	-	-	-	1	3	1	4

Figures

Figure 1

Distribution of CSE and CSE-like genes on *Populus* chromosomes. Regions that are assumed to correspond to homologous genome blocks are shaded gray and connected by lines. Schematic view of chromosome reorganization from the most recent whole-genome duplication in *Populus* was adapted from Tuskan et al. (2006), based on duplication coordinates from the *Populus* genome assembly version 3.0. The position of genes is indicated with an arrowhead, and positional paralogs relative to the duplicated blocks are connected by dashed lines.

Figure 2

Collinear analysis of homologous segments between *Populus* and *Amborella* Green arrow boxes present the CSE and CSE homologs; arrow boxes filled with other colors indicate other homologous genes (orthologs or paralogs). The symbols of CSE genes and the chromosomes they located were marked on right. Especially Poptr (chr 04 +) means no CSE homologs were found on this homologous segment (positive strand) on *Populus* chromosome 04.

AtCSE1	MPSAEASSNSAFATPPFPFNFGIMPEEEVYTSQVNRKSYFETENK.....PQCQLILD.GEIK.....GITY	68
PoptrCSE1	MPEAQQAAPFNFGIMPEEEVYASQVNTQSHFETPNK.....PQCQLILD.KRVK.....ATY	60
PoptrCSE2	MSSEIQQTFPFPNFGIMPEEEVYASQVNTQSHFETPNK.....PQCQLILD.KRVK.....ATY	60
PoptrCSE3	MPEQVAFANQTFPFPNFGIMPEEEVYASQVNTQSHFETPNK.....PQCQLILD.KRVK.....ATY	60
PoptrCSE4	MCMVVCQGVIRSS.....KQVQVPCRLISG.SSP.R...ALVFL	38
PoptrCSE5	MAREIGNVMYCEFLVNS.....KRLKPCQMIITN.KE.P.K...ALVEI	41
PoptrCSE6	MAREIGNVMYCEFLVNS.....KRLKPCQMIITN.KE.P.K...ALVEI	41
PoptrCSE7	MAREIGNVMYCEFLVNS.....KRLKPCQMIITN.KE.P.K...ALVEI	41
PoptrCSE8	MLSHFQANEKSPVGLTREETFYKHCILQESFNFKMKKQCFQFAPCDFTS.Q.L.L...KQIA	63
PoptrCSE9	RRALAEQIMTGNEDCRNNTFFLVANNAHVRSLIT.GL.L.CDFGIL	132
PoptrCSE10	RRRAREAFDQLGIDHILFKTPCDGLKMHSEYVNS.....KLELFRKSLKRS.SS.P.K...AVVCF	88
PoptrCSE11	MVVQFVAEANDCSFGILSPSEFYAKHCQVTHSEYIINSGLKPCQWTLFFETR.I.I...GCA	62
PoptrCSE12	RRRVSRFVQVQCLIDELFMFTFGC.KTGLSEVNS.....KLELFRKSLKRS.SS.P.K...AVVCF	148
PoptrCSE13	MVVQFVAEANDCSFGILSPSEFYAKHCQVTHSEYIINSGLKPCQWTLFFETR.I.I...GCA	63
PoptrCSE14	MENKQMFQCSFREDTS.TS.Q.P...KQIA	27
PoptrCSE15	VHRTKLIVATIVKWRAGGGGVAVVDFVGGRRAIKRVLLGDDTIVREFSLFVIAISDNPCQSTYS.VK.I.R...GLVL	170
PoptrCSE16	VHRTKLIVATIVKWRAGGGGVAVVDFVGGRRAIKRVLLGDDTIVREFSLFVIAISDNPCQSTYS.VK.I.R...GLVL	152
PoptrCSE17	RRALAEQIMTGNEDCRNNTFFLVANNAHVRSLIT.GL.L.CDFGIL	128
PoptrCSE18	MLTWVVLSEYVNS.....KLELFRKSLKRS.SS.P.K...AVVCF	37
AtCSE1	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	154
PoptrCSE1	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	146
PoptrCSE2	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	146
PoptrCSE3	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	149
PoptrCSE4	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	123
PoptrCSE5	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	126
PoptrCSE6	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	126
PoptrCSE7	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	124
PoptrCSE8	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	147
PoptrCSE9	I.MEL.NMGRVYDFKEMNANRNVYVGNHIGHSDD...MHNYGSGWLFQKICISFANMNAIFACILLESRSDD	214
PoptrCSE10	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	173
PoptrCSE11	VHRTKLIVATIVKWRAGGGGVAVVDFVGGRRAIKRVLLGDDTIVREFSLFVIAISDNPCQSTYS.VK.I.R...GLVL	150
PoptrCSE12	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	233
PoptrCSE13	VHRTKLIVATIVKWRAGGGGVAVVDFVGGRRAIKRVLLGDDTIVREFSLFVIAISDNPCQSTYS.VK.I.R...GLVL	151
PoptrCSE14	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	72
PoptrCSE15	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	232
PoptrCSE16	MHNYGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	232
PoptrCSE17	I.MEL.NMGRVYDFKEMNANRNVYVGNHIGHSDD...MHNYGSGWLFQKICISFANMNAIFACILLESRSDD	210
PoptrCSE18	C.HGYAMGSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	109
AtCSE1	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	241
PoptrCSE1	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	233
PoptrCSE2	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	233
PoptrCSE3	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	235
PoptrCSE4	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	209
PoptrCSE5	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	212
PoptrCSE6	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	212
PoptrCSE7	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	210
PoptrCSE8	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	234
PoptrCSE9	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	299
PoptrCSE10	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	259
PoptrCSE11	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	237
PoptrCSE12	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	319
PoptrCSE13	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	238
PoptrCSE14	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	243
PoptrCSE15	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	337
PoptrCSE16	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	317
PoptrCSE17	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	295
PoptrCSE18	YITLQ.RGAWGLIINGAMCSAKFEPWLEHLFVVA...VPTSVITRGSIFELSFHEKRLGCAQGRVTR.FRAAAYE	161
AtCSE1	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	331
PoptrCSE1	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	323
PoptrCSE2	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	323
PoptrCSE3	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	323
PoptrCSE4	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	328
PoptrCSE5	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	301
PoptrCSE6	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	301
PoptrCSE7	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	300
PoptrCSE8	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	323
PoptrCSE9	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	346
PoptrCSE10	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	386
PoptrCSE11	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	324
PoptrCSE12	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	405
PoptrCSE13	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	323
PoptrCSE14	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	232
PoptrCSE15	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	417
PoptrCSE16	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	397
PoptrCSE17	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	377
PoptrCSE18	INYPQSGWLFQKICISFANMNAIFACILLESRSDD...RCMHGMDHAAATLS.FKHVRCSDPYKIDAFIENKLVIT	193

Figure 3

Alignment of eighteen putative PoptrCSE sequences with AtCSE1 (At1g52760). Conserved and functional motifs are indicated in black boxes, including an acyltransferase motif (HX4D) and two lipase/hydrolase motifs (GXSGX). Key amino acids involved in enzymatic activity are also marked with black stars (acyltransferase activity) or circles (catalytic triad).

Figure 4

Phylogenetic tree of screened CSE and CSE-like protein sequences from *Populus* and other species. Eighteen PoptrCSE or PoptrCSE-like proteins are marked with red triangle filled circles. All 153 CSE and CSE-like proteins including AtMAGLs are listed in Supplementary Table S2.

Figure 5

Expression profile and co-expression network of CSE and CSE-like genes in poplar. **a**. Expression profile of PoptrCSE and PoptrCSE-like genes in various *Populus* tissues. Cluster analysis associated with the expression profile is also shown in this figure. Transcript levels are indicated by shading from dark red to white as shown in the legend. Tissues or specific parts of plants are indicated with the respective abbreviations: DP, developing phloem; DX, developing xylem; ML, mature leaf; SLp, shoot and leaf primordium. The support information including log2-transformed signal intensity and statistical test is available in Supplementary Table S5. **b**. Co-expression network of PoptrCSE and PoptrCSE-like genes with identified genes involved in lignin biosynthesis. PoptrCSE1, -2 and -12 were marked with red boxes. **c**. Similar expression pattern of PoptoCSE1, -2 and PoptoCAD, PoptoCCR7, Popto- CCoAOMT using Semi-qRT-PCR.

Figure 6

Subcellular localization of PoptoCSE1 and PoptoCSE2 in transformed tobacco leaves. The red arrow and white arrow indicate the plasma membrane and chloroplast, respectively. CP indicates chloroplast and PM indicates plasma membrane.

Figure 7

Enzymatic assay of PoptoCSE1 and PoptoCSE2. a-c: The decrease of caffeoyl shikimate and the increase of caffeic acid along with the reaction process catalyzed by PoptoCSE1 with different reaction time 5s(a), 80s(b) and 120s(c); d-e: The temperature (d) and pH (e) profiles for recombinant PoptoCSE1 and PoptoCSE2 activities. The mean value of three replicates were used to calculate the relative activity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.tif](#)
- [FigureS2mgalML.tif](#)
- [SupplementaryTableS1.xls](#)
- [SupplementaryTableS2PutativeCSE.xls](#)
- [SupplementaryTableS3.xls](#)
- [SupplementaryTableS4ConservedmotifsofputativeCSEs.xls](#)
- [SupplementaryTableS5.xls](#)
- [SupplementaryTableS6.xls](#)
- [SupplementaryTableS7Primersusedinthisstudy.xlsx](#)