Characterization and functional analysis of the Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (HCT) gene family in poplar

Nan Chao [1, 2, 3], Qi Qi [2, 4], Shuang Li [2], Brent Ruan [5], Xiang-Ning Jiang [2, 6], Ying Gai [Corresp. 2, 6]

1 School of Life Science, Tsinghua University, Beijing, China

2 College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing, China

3 School of Biotechnology, Jiangsu University of Science and Technology, Zhenjiang, China

4 College of Horticulture,, China Agricultural University, Beijing, China

5 Department of Agricultural and Biological Engineering, University of Illinois at Urbana Champaign, Urbana Champaign, USA

6 National Engineering Laboratory for Tree Breeding, the Tree and Ornamental Plant Breeding and Biotechnology Laboratory of Chinese Forestry Administration, Beijing, China

Corresponding Author: Ying Gai
Email address: gaiying@bjfu.edu.cn

Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (HCT) divides the mass flux to H, G and S units in monolignol biosynthesis and affects lignin content. Ten HCT homologs were identified in the *Populus trichocarpa* (Torr. & Gray) genome. Both genome duplication and tandem duplication resulted in the expansion of HCT orthologs in *Populus*. Comprehensive analysis including motif analysis, phylogenetic analysis, expression profiles and co-expression analysis revealed the divergence and putative function of these candidate *PoptrHCTs*. *PoptrHCT1* and *2* were identified as likely involved in lignin biosynthesis. *PoptrHCT9* and *10-* are likely to be involved in plant development and the response to cold stress. Similar functional divergence was also identified in *Populus tomentosa Carr.* Enzymatic assay of PtoHCT1 showed that PtoHCT1 was able to synthesize caffeoyl shikimate using caffeoyl-CoA and shikimic acid as substrates.

1 **Characterization and functional analysis of the Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl**

2 **transferase (HCT) gene family in poplar**

3 **Nan Chao**[1,2,4], **Qi Qi**[2,5], **Li Shuang**[2]**, Brent Ruan**[6]**, Xiang-Ning Jiang** [2,3] **and Ying Gai** [2,3†]

4 [1]School of Life Science, Tsinghua University,

5 [2]College of Biological Sciences and Biotechnology, Beijing Forestry University,

6 [3]National Engineering Laboratory for Tree Breeding, the Tree and Ornamental Plant Breeding and

7 Biotechnology Laboratory of Chinese Forestry Administration, Beijing 100083, China,

8 [4]School of Biotechnology, Jiangsu University of Science and Technology,

9 [5]College of Horticulture, China Agricultural University,

10 [6]Department of Agricultural and Biological Engineering, University of Illinois at Urbana Champaign

11 †To whom correspondence should be addressed: Tel (Fax): +86 10 6233 8063;

12 E-mail: gaiying@bjfu.edu.cn, Beijing Forestry University.

14 **The English in this document has been checked by at least two professional editors, both native speakers**

15 **of English, and the grammars were also checked using Grammarly.**

16   **Abstract:** Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (HCT) divides the mass flux to H,

17   G and S units in monolignol biosynthesis and affects lignin content. Ten HCT homologs were identified in the

18   *Populus trichocarpa* (Torr. & Gray) genome. Both genome duplication and tandem duplication resulted in the

19   expansion of HCT orthologs in *Populus*. Comprehensive analysis including motif analysis, phylogenetic

20   analysis, expression profiles and co-expression analysis revealed the divergence and putative function of these

21   candidate *PoptrHCTs*. *PoptrHCT1* and *2* were identified as likely involved in lignin biosynthesis. *PoptrHCT9*

22   and *10-* are likely to be involved in plant development and the response to cold stress. Similar functional

23   divergence was also identified in *Populus tomentosa Carr.* Enzymatic assay of PtoHCT1 showed that PtoHCT1

24   was able to synthesize caffeoyl shikimate using caffeoyl-CoA and shikimic acid as substrates.

25   **Key words**:Hydroxycinnamoyl- CoA: shikimate hydroxycinnamoyl transferase; enzymatic synthesis;

26   divergence; gene family; monolignol; *Populus*

27   **Introduction**

28   Primary walls and secondary walls protect plant cells and define the shapes of cells, tissues, organs

29   and ultimately the whole plant body (Zhong et al. 2019). Lignin is an important component for secondary

30   cell walls and is one of the most abundant components of biomass in plants (Boerjan et al. 2003; Tang & Tang

31   2014). Therefore, lignin plays a vital role in plant physiology. Owing to the recalcitrant chemical nature and the

32   complexity of lignin, lignin limits the conversion efficiency of lignocellulosic biomass to ethanol (Poovaiah et

33   al. 2014; Vanholme et al. 2010). Modifying trees to have less lignin or more-degradable lignin along with normal

34   growth, can reduce the high processing costs and carbon footprint of making paper, biofuels, and chemicals

35   (Ralph et al. 2019; Tang & Tang 2014; Wang et al. 2019; Xu & Li 2016; Zhao 2016).

36   The biosynthetic pathway for lignin has been studied extensively and the phenylpropane pathway which

37   begins with phenylalanine, is responsible for monolignol biosynthesis. (Boerjan et al. 2003; Karkonen &

38   Koutaniemi 2010; Maeda 2016; Ralph et al. 2019; Vanholme et al. 2010; Wang et al. 2019; Xu & Li 2016).

39   Monolignol is the general name for lignin building blocks. Our understanding of the monolignol biosynthetic

40   pathway has continued to grow, and now 11 enzyme families and 24 metabolites are associated with it

41   (Vanholme et al. 2019). Hydroxycinnamoyl- CoA: shikimate hydroxycinnamoyl transferase (HCT) is located at

42   a key point in the monolignol biosynthetic pathway and is conserved across all land plants. In conjunction with

43   C3H (*p*-coumarate 3-hydroxylase), HCT catalyzes two steps to direct the mass flux from the H monolignol to G

44   and S monolignols (Figure 1). HCT first catalyzes the coupling of *p*-coumaroyl-CoA with shikimate to produce

45   *p*-coumaroyl shikimate (Hoffmann et al. 2004; Hoffmann et al. 2003). Caffeoyl shikimate is generated by C3H

46   and is then transesterified by HCT to form caffeoyl-CoA. This reaction is probably reversible based on the

47   reported in vitro activity (Lepelley et al. 2007; Wang et al. 2014). Caffeoyl shikimate esterase (CSE), a new

48   member in monolignol biosynthesis pathway recently discovered in plants can hydrolyze caffeoyl shikimate to

49   release caffeate (Ha et al. 2016; Saleme et al. 2017; Vanholme et al. 2013; Vargas et al. 2016). Although down-

50   regulation of HCT expression improves forage digestibility and saccharification efficiency, it negatively affects

51   plant growth resulting in shorter plants (Li et al. 2010; Shadle et al. 2007).

52   HCT (GO:0102660) belongs to the BAHD acyltransferase family and is able to utilize many non-native

53   substrates. Some HCTs (also called HQT) can use quinate as a substrate in addition to shikimate (Eudes et al.

54   2016; Kim et al. 2013) for the biosynthesis of chlorogenic acid. As an acyl-CoA-dependent transferase, HCT is

55   capable of acylating a wide variety of acceptors, with some exhibiting broad substrate flexibility (Chiang et al.

56   2018; Eudes et al. 2016). Crystal structures of HCTs from different plants have been determined for both the

57   apo-form and complexed structure with diverse substrates allowing determination of active sites. For example,

58   the apo-form and ternary complex with *p*-coumaroyl-CoA and shikimate of SbHCT from *Sorghum bicolor* (L.)

59   Moench revealed the catalytic mechanism of HCT (Walker et al. 2013). Structures of AtHCT from *Arabidopsis*

60   *thaliana* (L.), CbHCT from *Coleus blumei* Benth, CcHCT from *Coffea canephora* Pierre ex Froehn and SmHCT

61   from *Selaginella moellendorffii* Hieron. have also been reported (Chiang et al. 2018; Lallemand et al. 2012;

62   Levsh et al. 2016).

63   Similar to other key genes involved in monolignol biosynthesis, HCT is found as a gene family in many

64   species in plant kingdom (Carocha et al. 2015; Ferreira et al. 2019; Ma et al. 2017; Raes et al. 2003; Zhang et al.

65   2018). The structural information and the proposed active sites of HCT, can help us to distinguish bona fide

66   HCT utilizing shikimate as an acceptor and involved in monolignol biosynthesis in plants. In this study, we used

67   genome-wide screening to identify 10 *HCT* homologs in *Populus trichocarpa*. Further motif and active site

68   analysis showed the divergence of *PoptrHCT*s. Expression profiles and co-expression network analysis

69   identified *PoptrHCT1* and *2* as the lignin-related HCTs. Finally, we cloned and characterized the catalytic

70   activity of PtoHCT1 from *Populus tomentosa in vitro,* which generated caffeoyl shikimate. PtoHCT1 could be

71   used as the target gene for genetic modification to alter lignin content and composition.

72   **Materials and Methods**

73   **Materials**

74   Leaves of six–-year-old *Populus tomentosa* 741 were collected from Hebei, China (Hu et al. 2019; Tian et

75   al. 2013). Samples were immediately frozen in liquid nitrogen and then stored at -80 ^{o}C until use.

76   **Genome-wide identification of HCT gene family members**

77   To identify the HCT sequences in *Populus,* we first built a hidden Markov model (HMM) using reported

78   HCT and HQT sequences. HMMsearch using Hmmer 3.0 software against the proteome data of *Populus*

79   *trichocarpa* was performed based on the HMM model (Eddy 2010). The cutoff for PoptrHCT homolog screening

80   was an E-value (<E–100) of both the domain and full sequence and scores of full sequences (>400)

81   (Supplementary Table S1). The stable gene ID and symbols for HCTs reported in a previous study were also

82   marked in Supplementary Table S1 Table S2. The sequences used for building the HMM model are shown in

83   Supplementary TableS2.

84   **Distribution of HCT genes and HCT orthologs on *Populus* chromosomes**

85   Ten candidate HCT and HCT orthologs were located on chromosomes in specific duplicated blocks which

86   were determined based on the *Populus* genome and the WGDotplot in the PLAZA platform (Proost et al. 2009).

87   **HCT sequence alignment and phylogenetic analysis**

88   Alignment of PoptrHCTs and SbHCT, AtHCT were performed using DNAman 8.0 (Lynnon BioSoft) with

89   default parameters. A phylogenetic tree was obtained using Mega 7.0 with the maximum-likelihood method

90   (Kumar et al. 2016; Tamura et al. 2011). The phylogenetic tree was assessed by bootstrapping using 1000

91   bootstrap replicates and marked above nodes only if greater than 50. The JTT substitution model and G+I rates

92   among sites model were selected as parameters for building the tree. The putative HCT sequences are listed in

93   Supplementary Table S2.

94   **HCT expression profiles in *P. trichocarpa* and *P. tomentosa***

95   We obtained gene expression profiles for various tissues in *P. trichocarpa* using the GEO database with the

96   accession number GSE30507. In addition, RNA-seq dataset GSE78953 including the transcriptome of various

97   monolignol biosynthesis related mutants in *P. trichocarpa,* was used for co-expression analysis to explore the

98   functions of the *PoptrHCT* orthologs. We also examined the expression profiles of *PtoHCT* orthologs in *P.*

99   *tomentosa* in different seasons (Spring, Summer, Fall and Winter) and organs or tissues (roots, buds, phloem

100   and xylem) using our microarray dataset (accession number: GSE56023) (Chao et al. 2014b). The corresponding

101   PtoHCTs were identified using PoptrHCTs as queries by local blastn against the probe sequences database

102   (Christiam et al. 2009) TBtools v0.6652 and Cytoscape 3.4 were used to visualize the HCT expression profile

103   or co-expression network (Chen et al. 2018; Shannon et al. 2003) (Supplementary Table S3).

104   **Cloning and purification of recombinant HCT from *P. tomentosa***

105   Isolation of RNA and cDNA synthesis have been described in a previous study (Chao et al. 2014a) .We

106   cloned the homologous *HCT1* from *P. tomentosa* based on the sequence information from *P. trichocarpa*

107   (GenBank accession number: KT021003). Primer pair used for PCR amplification of *PtoHCT1* is as follow:

108   forward, 5'-CGATAAATAGAGCATTAGCACGGGG-3'; and reverse, 5'-ATAG CCTCGGCTCATTCTTT-

109   3'. PCR products were purified and cloned into the pMD18-T vector (Takara Dalian), propagated in *Escherichia*

110   *coli DH5α* and inserts were confirmed by sequencing. *PtoHCT1* was constructed with pET28a (Novagen)

111   through a digestion-ligation way using restriction enzymes BamHI, HindIII and T4 ligation (Takara, Dalian).

112   pET28a -PtoHCT1was then transformed into *E.coli* BL21(DE3). To induce expression, Isopropy-β-D-

113   thiogalactoside (IPTG)was added to a final concentration of 0.8mM and incubation was continued at 28 °C for

114   four hours. Cells were collected by centrifugation at 4000g and 4℃ for 15min. The pellets were resuspended in

115   lysis buffer (50 mM NaH_2PO4, 300 mM NaCl with 10 mM imidazole, pH 8.0) and then disrupted by sonication.

116   After centrifugation at 12,000 g and 4 ℃ for 30 min, the lysates were mixed with pretreated 1 ml Ni-NTA agarose

117   (Qiagen Shanghai, China)) After washing using lysis buffer supplemented with 20mM imidazole, the His-tagged

118   PtoHCT1was eluted with 100mM imidazole in lysis buffer.

119   **Catalytic activity of recombinant PtoHCT1**

120   Caffeoyl-CoA was chemically synthesized as reported (Chao et al. 2017). We determined the activity of

121   recombinant PtoHCT1 by synthesis of caffeoyl shikimate using caffeoyl-CoA and shikimic acid as substrates.

122   The reaction was performed according to Cesarino *et al.*(2013) and Luis *et al.*(2014). Briefly, total 40μl standard

123   reaction mix contained 100 mM Tris-HCl pH 7, 1mM DTT, 100 μM caffeoyl-CoA, 100μM shikimic acid and

124   10 µg purified recombinant HCT protein. The reaction was initiated by adding the HCT proteins or the same

125   amount of boiled protein as negative control. After Incubating at 30 °C for 30 min, the reaction was terminated

126   by boiling the samples for 5 min. Flow for HPLC analysis was 0.1 mL/min in solvent A (acetonitrile) and solvent

127   B (0.01% formic acid in water). The gradient was 0% A to 35% in B for 0 to 24 min, 35% A in B to 100% B for

128   24 to 27min, 100% B for 27 to 32min, 100% A to 100% B for 32 to 35 min, and 100% B for 35 to 55 min. The

129   parameters used for MS analysis was sheath gas (nitrogen) flow rate, 40 arb; aux/sweep gas (nitrogen) flow rate,

130   10 arb; spray voltage,4.5 kV; capillary temperature, 320 °C. Optimized detailed parameters for dissociation of

131   parent ions into product ions for each compound were provided in Supplementary Table S4.

132   **Structure modeling of PtoHCT1**

133   The crystal structure of AtHCT (accession number 5KJT) (Levsh et al. 2016) was obtained from the Protein

134   Data Bank to build a homolog model for PtoHCT (https://www.rcsb.org). Molecular docking was performed

135   using CDOCKER assembled in Discovery Studio 4.5. Visualization of the active sites and 3-D structures were

136   generated by Discovery Studio 4.5.

137   **Results**

138   **Genome-wide identification and distribution of HCT orthologs in *Populus***

139   Ten *PoptrHCT* homologs were found based on HMMsearch against the *Populus* genome (Supplementary Table

140   S1). These *HCT* candidate genes are located on six different chromosomes. Among the 10 *PoptrHCT* orthologs,

141   *PoptrHCT3, 4* and *5* were located on chromosome V, and *PoptrHCT7, 8, 9,* and *10,* were on chromosome XVIII,

142   representing two clusters respectively (Figure 2A). Tandem duplication is likely to be responsible for the

143   formation of HCT homolog clusters. *Ks* (substitution per synonymous site) value distributions can be used for

144   revealing whole genome duplication (WGD) events (Jiao et al. 2011; Tang et al. 2010). *PoptrHCT1* and

145   *PoptrHCT2* formed a homolog duplicate pair with *Ks* value 0.2174 and were located at corresponding

146   homologous duplicated blocks, as the result of whole genome duplication. The organization of the PoptrHCT

147   orthologs indicates that both genome duplication and tandem duplication played roles in the formation of the

148   HCT family.

149   **Alignment and phylogenetic analysis of HCT orthologs**

150   Putative protein sequences of PoptrHCT orthologs and crystal structures of two shikimate-specific HCTs

151   (AtHCT and SbHCT) and LeHQT (*Lycopersicon esculentum* Mill.) were aligned. Characteristic of the BADH

152   superfamily, two motifs HXXXD(G) and DFGWG were conserved in AtHCT , SbHCT and all PoptrHCT

153   orthologs (except PoptrHCT6) (Figure 3) (D'Auria 2006). Based on previous studies of the structure of HCTs

154   including site-directed mutagenesis, molecular docking and crystallographic analyses we summarized the active

155   sites of HCTs (**Table1**) and marked these active sites in Figure 3. Active sites for the carbonyl group of the *p*-

156   coumaroyl moiety binding and the catalysis related sites of LeHQT correspond with HCTs (red full circles)

157   while divergence is obvious in terms of active sites for shikimate binding (red full stars). PoptrHCT1 and

158   PoptrHCT2 showed conservation at these active sites and kept correspondence with AtHCT and SbHCT, while

159   PoptrHCT3-10 showed poor conservation at these key sites. Thus while the ten candidate PoptrHCTs mostly

160   belong to the BADH superfamily, only PoptrHCT1 and PoptrHCT2 appear to be associated with monolignol

161   biosynthesis. Phylogenetic analysis showed PoptrHCT1 and PoptrHCT2 grouped with Group I HCTs, which

162   transfer hydroxycinnamates to shikimate and have been implicated in monolignol biosynthesis, strongly

163   suggesting this role for PoptrHCT1 and 2 as well (Figure 2B). Other PoptrHCTs (3-10) clustered with HQTs

164   and other HCT-like as Group II and especially, PoptrHCT6 without DFGWG seem unlikely to be shikimate-

165   specific transferases involved in monolignol biosynthesis. Thus these Group II. PoptrHCTs might have different

166   catalytic activity (e.g., utilize acceptors other than shikimic acid) and are likely to play different roles in plants.

167   **Expression analysis of HCT homolog genes**

168   Based on the microarray analysis of seven different tissues and organs in *P. trichocarpa*, *PoptrHCT1* and *2*

169   showed expression preference in developing xylem (DX) and mature xylem (MX). PoptrHCT1 showed high

170   expression levels in all detected tissues and organs. PoptrHCT9 and 10 showed expression preference in

171   developing tissues including developing phloem, developing xylem, cambium (C) and shoots and leaf

172   primordium (SLp) **(Figure 4A)**. Co-expression network analysis shows that *PoptrHCT1* and *2* have significant

173   correlations with genes involved in lignin biosynthesis (Figure 4B)**.** Monolignol biosynthesis related

174   transcription factors also showed co-expression with *PoptrHCT1* and *2*. The similar expression patterns were

175   also found in *P. tomentosa* Carr. *(Pto)*. *PtoHCT1* has high expression levels in almost all tissues and organs year-

176   round while *PtoHCT9* and *10* showed preference in buds and phloem especially in winter, which indicates that

177   these two genes could be involved in development of dormancy and response to cold stress. The differential

178   expression of the *PoptrHCT* orthologs further supports that *HCT1* plays a major role in monolignol biosynthesis.

179   **Catalytic activity and structure comparison of PtoHCT1**

180   PtoHCT1 protein expressed in *E. coli* was purified for enzymatic assays. We monitored PtoHCT1 reactions using

181   HPLC-MS and found that PtoHCT1can utilize caffeoyl-CoA and shikimic acid **(Figure 5)**. After the initiation

182   of the reaction by adding PtoHCT1, the accumulation of caffeoyl shikimate and decrease of caffeoyl-CoA and

183   shikimic acid were observed within 2 minutes. We built a homology model for PtoHCT1 to explore the structure

184   of PtoHCT1 using the crystal structure of AtHCT (accession number 5KJT) as template. The main-chain root-

185   mean-square deviation (RMSD) is 0.224 Å indicating the high structure similarity of PtoHCT1 and AtHCT

186   (**Figure 6A**). According to the summarized active sites (Table 1), we found these conserved sites around the

187   catalytic cleft, and then we successfully docked PtoHCT1 with the substrate caffeoyl-CoA, which further

188   provided the structural evidence for PtoHCT1 catalyzing caffeoyl-CoA (**Figure 6B**).

189   **Discussion**

190   HCT regulates the flux at a key point in monolignol biosynthesis and has been studied in many plants

191   (Hoffmann et al. 2004; Hoffmann et al. 2003; Shadle et al. 2007; Sun et al. 2018; Wagner et al. 2007). HCT also

192   exists as a gene family in the land plant kingdom similar to other key genes involved in monolignol biosynthesis.

193   We focused on the HCT gene family in this study and provided a systematic analysis of the HCT genes in poplar,

194   and identified two lignin-related HCTs (*HCT1*and *HCT2*), which exist as a homolog pair located at a duplication

195   block on chromosome I and III respectively (Figure 2A). Ks analysis indicated that the HCT gene pair

196   (*PoptrHCT1* and *PoptrHCT2*) resulted from a recent genome duplication (Ks value 0.2174). Two HCT homolog

197   clusters resulting from tandem duplication were also identified. Both genome duplication and tandem duplication

198   provide raw genetic material for neo-function as a driving force in plant evolution and are responsible for the

199   expansion of HCT orthologs (Zhang 2003).

200   Based on our systematic analysis, *PoptrHCT1* and *2* are involved in monolignol biosynthesis. Both

201   *PoptrHCT1* and *2* showed expression preference in xylem and co-expression with other monolignol related

202   genes. *HCT1* (*PoptrHCT1* and *PtoHCT1*) had high expression levels in different tissues. PtoHCT1 showed

203   catalytic activity for caffeoyl-CoA and shikimic acid. These results further validate *HCT1* as the dominant HCT

204   in monolignol biosynthesis, while *HCT9* and *10* (*PoptrHCT9*, 10 and *PtoHCT9, 10*) showed expression

205 preference in developing tissues. *PtoHCT9* and *10* were active in winter, which suggested a function in plant

206 development and response to cold stress. PtoHCT1 utilizes caffeoyl-CoA and shikimic acid to generate caffeoyl

207 shikimate which can be used as substrate for CSE, a new annotated enzyme involved in monolignol biosynthesis

208 (Ha et al. 2016; Saleme et al. 2017; Vanholme et al. 2013).

209 **Conclusion**

210 In summary, we identified ten HCT homologs and proposed the important roles of both genome duplication

211 and tandem duplication in the expansion of HCT orthologs in *Populous*. Two HCTs likely involved in

212 monolignol biosynthesis in *Populus* were identified based on phylogenetic analysis and expression profile

213 analysis. Enzymatic assay of PtoHCT1 showed that PtoHCT1 was able to synthetize caffeoyl shikimate using

214 caffeoyl-CoA and shikimic acid as substrates. In addition, other *PoptrHCT* orthologs showed divergence in

215 reported active sites and different expression pattern. *HCT9* and *10* (*PoptrHCT9*, 10 and *PtoHCT9, 10*) showed

216 preferential expression in developing tissues and were active in winter. Further studies should help to reveal the

217 functions of the other HCT orthologs.

218 Supplementary Table S1 Genome-wide screening *PoptrHCT* genes based on HMM models;

219 Supplementary Table S2 Putative HCTs used for phylogenetic analysis;

220 Supplementary Table S3 Expression profiles and co-expression for HCT homologs.

221 Supplementary Table S4 Conditions used for HPLC-MS to identify caffeic acid, shikimate and caffeoyl

222 shikimate

223 Supplementary File 5: Sequences used for alignment analysis.

224 **Acknowledgements**

225 This work was jointly supported by the Beijing Higher Education Young Elite Teacher Project [YETP0755

226 granted to Dr. Ying Gai], the National Natural Science Foundation [NSF 31300498 to Ying Gai].

227 **Author Contributions**

228 Ying Gai and Xiang-Ning Jiang guided the work and provided advice; Qi Qi, Shuang Li and Nan Chao performed

229 the experiments; Brent Ruan and Nan Chao read and revised the draft; Nan Chao analyzed the data and wrote

230 the paper.

231 **Conflict of Interest**

232   The authors declare that the research was conducted in the absence of any commercial or financial

233   relationships that could be construed as a potential conflict of interest.

234   **Reference**

235   Boerjan W, Ralph J, and Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* 54:519-546.
236   10.1146/annurev.arplant.54.031902.134938

237   Carocha V, Soler M, Hefer C, Cassan-Wang H, Fevereiro P, Myburg AA, Paiva JA, and Grima-Pettenati J. 2015.
238   Genome-wide analysis of the lignin toolbox of Eucalyptus grandis. *New Phytologist* 206:1297-1313.
239   10.1111/nph.13313

240   Chao N, Li N, Qi Q, Li S, Lv T, Jiang X-N, and Gai Y. 2017. Characterization of the cinnamoyl-CoA reductase (CCR) gene
241   family in Populus tomentosa reveals the enzymatic active sites and evolution of CCR. *Planta* 245:61-75.

242   Chao N, Liu S-X, Liu B-M, Li N, Jiang X-N, and Gai Y. 2014a. Molecular cloning and functional analysis of nine cinnamyl
243   alcohol dehydrogenase family members in Populus tomentosa. *Planta* 240:1097-1112.

244   Chao N, Liu SX, Liu BM, Li N, Jiang XN, and Gai Y. 2014b. Molecular cloning and functional analysis of nine cinnamyl
245   alcohol dehydrogenase family members in Populus tomentosa. *Planta* 240:1097-1112. 10.1007/s00425-
246   014-2128-9

247   Chen C, Xia R, Chen H, and He Y. 2018. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools
248   with a user-friendly interface.

249   Chiang YC, Levsh O, Lam CK, Weng JK, and Wang Y. 2018. Structural and dynamic basis of substrate permissiveness
250   in hydroxycinnamoyltransferase (HCT). *PLoS Comput Biol* 14:e1006511. 10.1371/journal.pcbi.1006511

251   Christiam, Camacho, George, Coulouris, Vahram, Avagyan, Ning, Ma, Jason, and Papadopoulos. 2009. BLAST+:
252   architecture and applications. *BMC Bioinformatics*.

253   D'Auria JC. 2006. Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology* 9:331-340.

254   Eddy S. 2010. HMMER3: a new generation of sequence homology search software. *URL: http://hmmer janelia Org*.

255   Eudes A, Pereira JH, Yogiswara S, Wang G, Benites VT, Baidoo EE, Lee TS, Adams PD, Keasling JD, and Loqué D. 2016.
256   Exploiting the Substrate Promiscuity of Hydroxycinnamoyl-CoA: Shikimate Hydroxycinnamoyl Transferase
257   to Reduce Lignin. *Plant and Cell Physiology* 57:568-579.

258   Ferreira SS, Simoes MS, Carvalho GG, De Lima LGA, Svartman RMDA, and Cesarino I. 2019. The lignin toolbox of the
259   model grass Setaria viridis. *Plant Molecular Biology*:1-21.

260   Ha CM, Escamilla-Trevino L, Yarce JC, Kim H, Ralph J, Chen F, and Dixon RA. 2016. An essential role of caffeoyl
261   shikimate esterase in monolignol biosynthesis in Medicago truncatula. *Plant Journal* 86:363-375.
262   10.1111/tpj.13177

263   Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, and Legrand M. 2004. Silencing of
264   hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid
265   biosynthesis. *The Plant Cell* 16:1446-1465.

266   Hoffmann L, Maury S, Martz F, Geoffroy P, and Legrand M. 2003. Purification, cloning, and properties of an
267   acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism.
268   *Journal of Biological Chemistry* 278:95-103.

269   Hu JQ, Qi Q, Zhao YL, Tian XM, Lu H, Gai Y, and Jiang XN. 2019. Unraveling the impact of Pto4CL1 regulation on the
270   cell wall components and wood properties of perennial transgenic Populus tomentosa. *Plant Physiology*
271   *and Biochemistry* 139:672-680. 10.1016/j.plaphy.2019.03.035

272 Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, and Soltis PS.
273   2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97-100.
274 Karkonen A, and Koutaniemi S. 2010. Lignin biosynthesis studies in plant tissue cultures. *J Integr Plant Biol* 52:176-
275   185. 10.1111/j.1744-7909.2010.00913.x
276 JIPB913 [pii]
277 Kim YB, Thwe AA, Kim YJ, Li X, Kim HH, Park PB, Suzuki T, Kim S-J, and Park SU. 2013. Characterization of Genes for a
278   Putative Hydroxycinnamoyl-coenzyme A Quinate Transferase and p-Coumarate 3′-Hydroxylase and
279   Chlorogenic Acid Accumulation in Tartary Buckwheat. *Journal of Agricultural and Food Chemistry* 61:4120-
280   4126.
281 Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger
282   Datasets. *Molecular Biology & Evolution* 33:1870.
283 Lallemand LA, Zubieta C, Lee SG, Wang Y, Acajjaoui S, Timmins J, Mcsweeney S, Jez JM, Mccarthy J, and Mccarthy
284   AA. 2012. A STRUCTURAL BASIS FOR THE BIOSYNTHESIS OF THE MAJOR CHLOROGENIC ACIDS FOUND IN
285   COFFEE. *Plant Physiology* 160:249-260.
286 Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, and McCarthy J. 2007. Chlorogenic acid synthesis in coffee:
287   An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes
288   during grain development in C. canephora. *Plant Science* 172:978-996.
289 Levsh O, Chiang Y, Tung CF, Noel JP, Wang Y, and Weng J. 2016. Dynamic Conformational States Dictate Selectivity
290   toward the Native Substrate in a Substrate-Permissive Acyltransferase. *Biochemistry* 55:6314-6326.
291 Li X, Bonawitz ND, Weng JK, and Chapple C. 2010. The growth reduction associated with repressed lignin biosynthesis
292   in Arabidopsis thaliana is independent of flavonoids. *Plant Cell* 22:1620-1632. 10.1105/tpc.110.074161
293 Ma C, Zhang H, Li J, Tao S, Qiao X, Korban SS, Zhang S, and Wu J. 2017. Genome-wide analysis and characterization
294   of molecular evolution of the HCT gene family in pear (Pyrus bretschneideri). *Plant Systematics and*
295   *Evolution* 303:71-90.
296 Maeda HA. 2016. Lignin biosynthesis: Tyrosine shortcut in grasses. *Nat Plants* 2:16080. 10.1038/nplants.2016.80
297 Poovaiah CR, Nageswara‐Rao M, Soneji JR, Baxter HL, and Stewart CN. 2014. Altered lignin biosynthesis using
298   biotechnology to improve lignocellulosic biofuel feedstocks. *Plant Biotechnology Journal*.
299 Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, and Vandepoele K. 2009. PLAZA: a comparative
300   genomics resource to study gene and genome evolution in plants. *Plant Cell* 21:3718-3731.
301   10.1105/tpc.109.071506
302 Raes J, Rohde A, Christensen JH, De Peer YV, and Boerjan W. 2003. Genome-Wide Characterization of the
303   Lignification Toolbox in Arabidopsis. *Plant Physiology* 133:1051-1071.
304 Ralph J, Lapierre C, and Boerjan W. 2019. Lignin structure and its engineering. *Current Opinion in Biotechnology*
305   56:240-249. 10.1016/j.copbio.2019.02.019
306 Saleme MLS, Cesarino I, Vargas L, Kim H, Vanholme R, Goeminne G, Van Acker R, Fonseca FCA, Pallidis A, Voorend
307   W, Junior JN, Padmakshan D, Van Doorsselaere J, Ralph J, and Boerjan W. 2017. Silencing CAFFEOYL
308   SHIKIMATE ESTERASE Affects Lignification and Improves Saccharification in Poplar. *Plant Physiology*
309   175:1040-1057. 10.1104/pp.17.00920
310 Shadle G, Chen F, Reddy MS, Jackson L, Nakashima J, and Dixon RA. 2007. Down-regulation of hydroxycinnamoyl
311   CoA: shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and
312   forage quality. *Phytochemistry* 68:1521-1529.

313   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape:
314   A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*
315   13:2498-2504.

316   Sun CH, Yang CY, and Tzen JTC. 2018. Molecular Identification and Characterization of Hydroxycinnamoyl Transferase
317   in Tea Plants (Camellia sinensis L.). *Int J Mol Sci* 19. 10.3390/ijms19123938

318   Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: molecular evolutionary genetics
319   analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular*
320   *biology and evolution* 28:2731-2739.

321   Tang H, Bowers JE, Wang X, and Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the
322   monocot lineage. *Proceedings of the National Academy of Sciences* 107:472-477.

323   Tang W, and Tang AY. 2014. Transgenic woody plants for biofuel. *Journal of Forestry Research* 25:225-236.

324   Tian X, Xie J, Zhao Y, Lu H, Liu S, Qu L, Li J, Gai Y, and Jiang X. 2013. Sense-, antisense- and RNAi-4CL1 regulate soluble
325   phenolic acids, cell wall components and growth in transgenic Populus tomentosa Carr. *Plant Physiology*
326   *and Biochemistry* 65:111-119. 10.1016/j.plaphy.2013.01.010

327   Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, and Araujo P. 2013. Caffeoyl
328   shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science* 341:1103-
329   1106.

330   Vanholme R, De Meester B, Ralph J, and Boerjan W. 2019. Lignin biosynthesis and its integration into metabolism.
331   *Current Opinion in Biotechnology* 56:230-239. 10.1016/j.copbio.2019.02.018

332   Vanholme R, Demedts B, Morreel K, Ralph J, and Boerjan W. 2010. Lignin biosynthesis and structure. *Plant Physiology*
333   153:895-905. 10.1104/pp.110.155119

334   pp.110.155119 [pii]

335   Vargas L, Cesarino I, Vanholme R, Voorend W, de Lyra Soriano Saleme M, Morreel K, and Boerjan W. 2016. Improving
336   total saccharification yield of Arabidopsis plants by vessel-specific complementation of caffeoyl shikimate
337   esterase (cse) mutants. *Biotechnol Biofuels* 9:139. 10.1186/s13068-016-0551-9

338   Wagner A, Ralph J, Akiyama T, Flint H, Phillips L, Torr K, Nanayakkara B, and Te Kiri L. 2007. Exploring lignification in
339   conifers by silencing hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyltransferase in Pinus radiata.
340   *Proceedings of the National Academy of Sciences* 104:11856-11861.

341   Walker AM, Hayes RP, Youn B, Vermerris W, Sattler SE, and Kang C. 2013. Elucidation of the structure and reaction
342   mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other coenzyme A-
343   dependent transferases and synthases. *Plant Physiology* 162:640-651.

344   Wang JP, Matthews ML, Naik PP, Williams CM, Ducoste JJ, Sederoff RR, and Chiang VL. 2019. Flux modeling for
345   monolignol biosynthesis. *Current Opinion in Biotechnology* 56:187-192. 10.1016/j.copbio.2018.12.003

346   Wang JP, Naik PP, Chen HC, Shi R, Lin CY, Liu J, Shuford CM, Li Q, Sun YH, Tunlaya-Anukit S, Williams CM, Muddiman
347   DC, Ducoste JJ, Sederoff RR, and Chiang VL. 2014. Complete proteomic-based enzyme reaction and
348   inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in
349   Populus trichocarpa. *Plant Cell* 26:894-914. 10.1105/tpc.113.120881

350   Xu P, and Li L. 2016. Monolignol Biosynthesis and Regulation in Grasses. *Recent Advances in Polyphenol Research*
351   5:108.

352   Zhang J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 18:292-298.

353   Zhang J, Yang Y, Zheng K, Xie M, Feng K, Jawdy SS, Gunter LE, Ranjan P, Singan VR, Engle N, Lindquist E, Barry K,

354   Schmutz J, Zhao N, Tschaplinski TJ, LeBoldus J, Tuskan GA, Chen JG, and Muchero W. 2018. Genome-wide

355   association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in

356   caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in Populus.

357   *New Phytologist* 220:502-516. 10.1111/nph.15297

358 Zhao Q. 2016. Lignification: flexibility, biosynthesis and regulation. *Trends in Plant Science* 21:713-721.

359 Zhong R, Cui D, and Ye ZH. 2019. Secondary cell wall biosynthesis. *New Phytologist* 221:1703-1723.

360   10.1111/nph.15537

361

Figure 1

Schematic diagram of reaction catalyzed by HCT in monolignol biosynthesis pathway

R=Shikimate. 4CL, 4-coumarate-CoA ligase; C3H, p-coumarate 3-hydroxylase; HCT, Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase. Compounds in red shadow are precursors for H units and in green shadow are for G and S units.

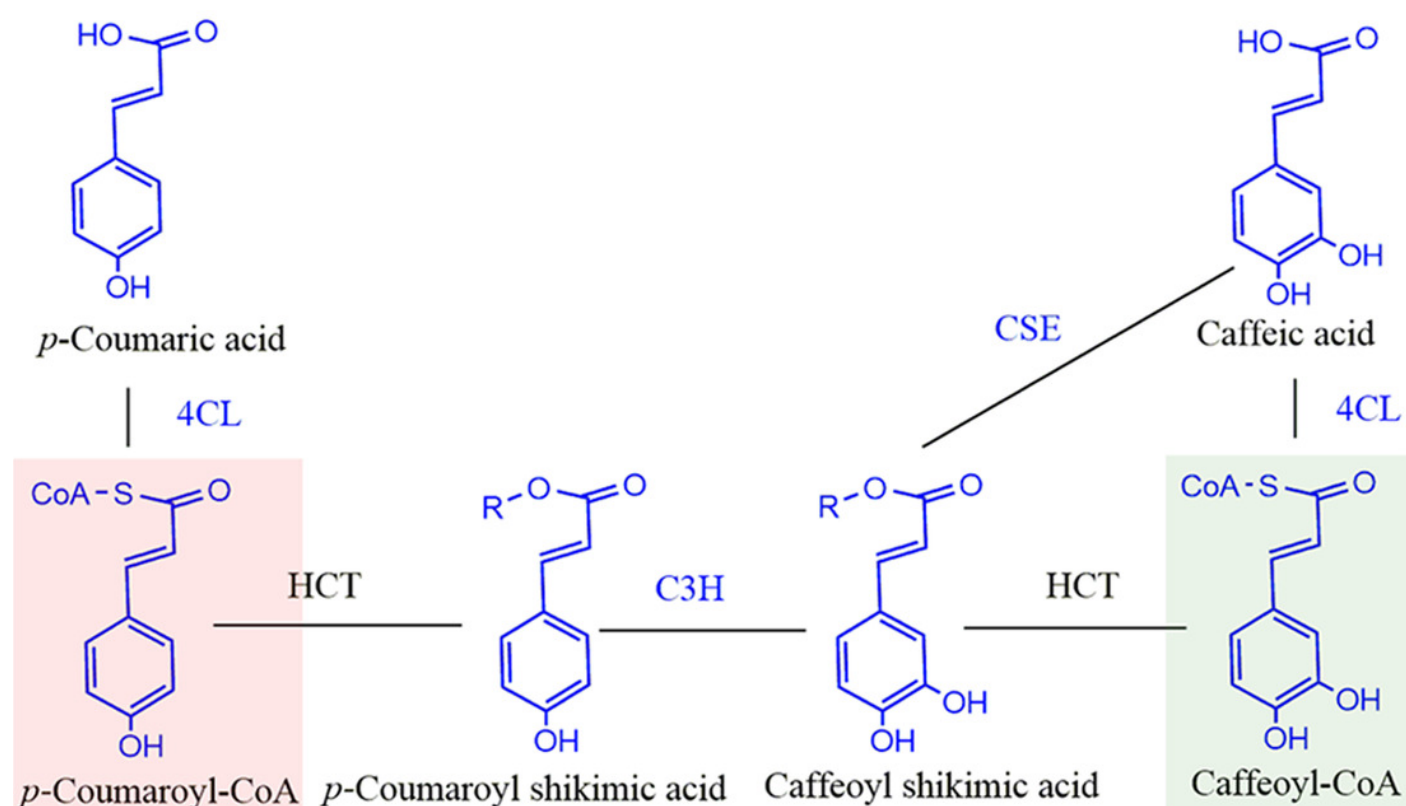



Figure 2

PoptrHCT orthologs organization and phylogenetic analysis

(A). Organization of HCT orthologs on *Populus* chromosomes. Regions that are assumed to correspond to homologous genome blocks are shaded gray and connected by lines. The position of genes is indicated with an arrowhead. (B). Phylogenetic analysis of HCT homologs from *Populus trichocarpa* and other plant species. The PoptrHCT1 and 2 were marked with full black triangle. Two groups for HCT orthologs were shown and HCTs in Group I are likely to transfer hydroxycinnamates to shikimate and have been implicated in monolignol biosynthesis. The scale bar indicates 0.5 amino acid substitutions per site in given length. The accession numbers of sequences used are as followed: *Arabidopsis thialiana* AtHCT (AT5G48930), AtHCTlike (AT4G29250); *Amborella trichopoda* AtrHCT1 (ATR_00137G00320), AtrHCT2 (ATR_00727G00010); *Cynara cardunculus* CcaHCT (DQ104740), *CcaHQT (ABK79690); Lycopersicon esculentum* LeHQT (AJ582652); *Larix kaempferi* LkaHCT (AHA44839); *Nicotiana tabacum* NtaHCT (Q8GSM7), NtaHQT(CAE46932); *Picea lauca* PglHCT (CZO01061061); *Populus trichocarpa* PoptrHCT1 (PT01G04290), PoptrHCT2 (PT03G18390), PoptrHCT3 (PT05G02800), PoptrHCT4 (PT05G02810), PoptrHCT5 (PT05G02840), PoptrHCT6 (PT18G03270), PoptrHCT7 (PT18G10470), PoptrHCT8 (PT18G10480), PoptrHCT9 (PT18G10540), PoptrHCT10 (PT18G10550); *Physcomitrella patens* PpHCT1 (PP00022G00830); *Panicum virgatum* PviHCT1a (JX845714), PviHCT2 (KC696573*); Sorghum bicolor SbHCT* (XP_002452435.1)."

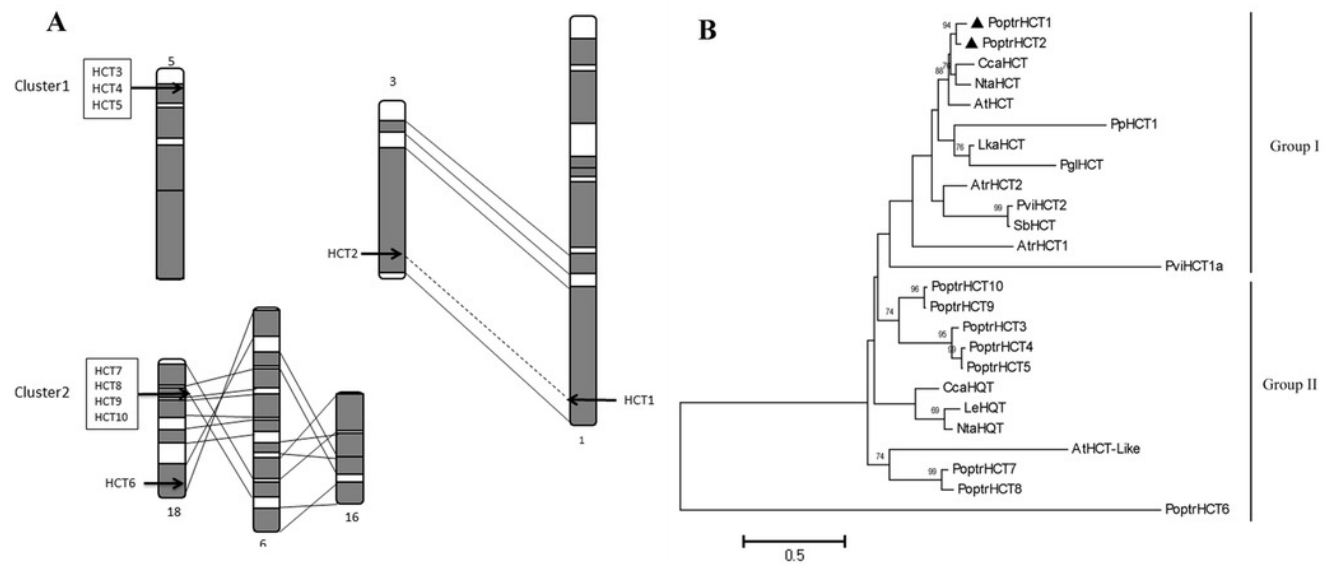



Figure 3

Alignment of PoptrHCT and PoptrHCT orthologs compared to shikimate-specific HCTs from Arabidopsis *and* sorghum *and HQT from tomato.*

Red full stars indicate shikimate binding sites, red full circles indicate carbonyl group of p-coumaroyl moiety binding sites, purple full triangle indicate carbonyl group of shikimate moiety binding sites and the blue full circles indicate sites involved in catalysis. Accessions are as in Figure 2. Detailed references are also available in Table 1

Figure 4

Expression profile and co-expression network of HCT orthologs in poplar.

(A). Expression profile of HCT orthologs in Populus. Tissues or specific parts of plants are indicated with the respective abbreviations: WS, whole stems; BM, Bark and mature phloem; C, cambium; DP, developing phloem; DX, developing xylem; ML, mature leaf; SLp, shoot and leaf primordium. (B) Expression profile of HCT orthologs in P. tomentosa. *(C) Co-expression network of PoptrHCT orthologs with identified genes involved in lignin biosynthesis. The support information is available in Table S3. Only nodes with Pearson correlation coefficients >0.9 were shown and considered as close co-expression.*

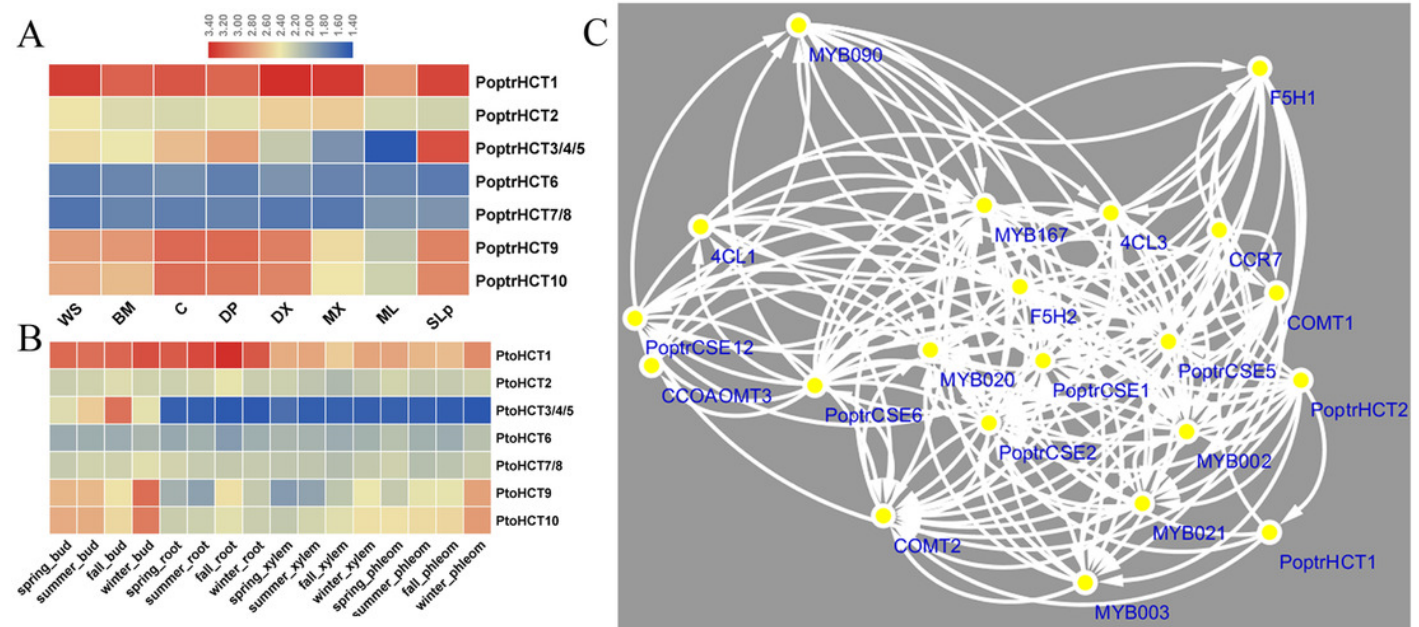



Figure 5

PtoHCT1 catalyzes enzymatic synthesis of caffeoyl shikimate.

LC separation of reactions with MS detection (selected ion signals) (A) at initiation of the reaction (B) after 80s or (C) after 120s

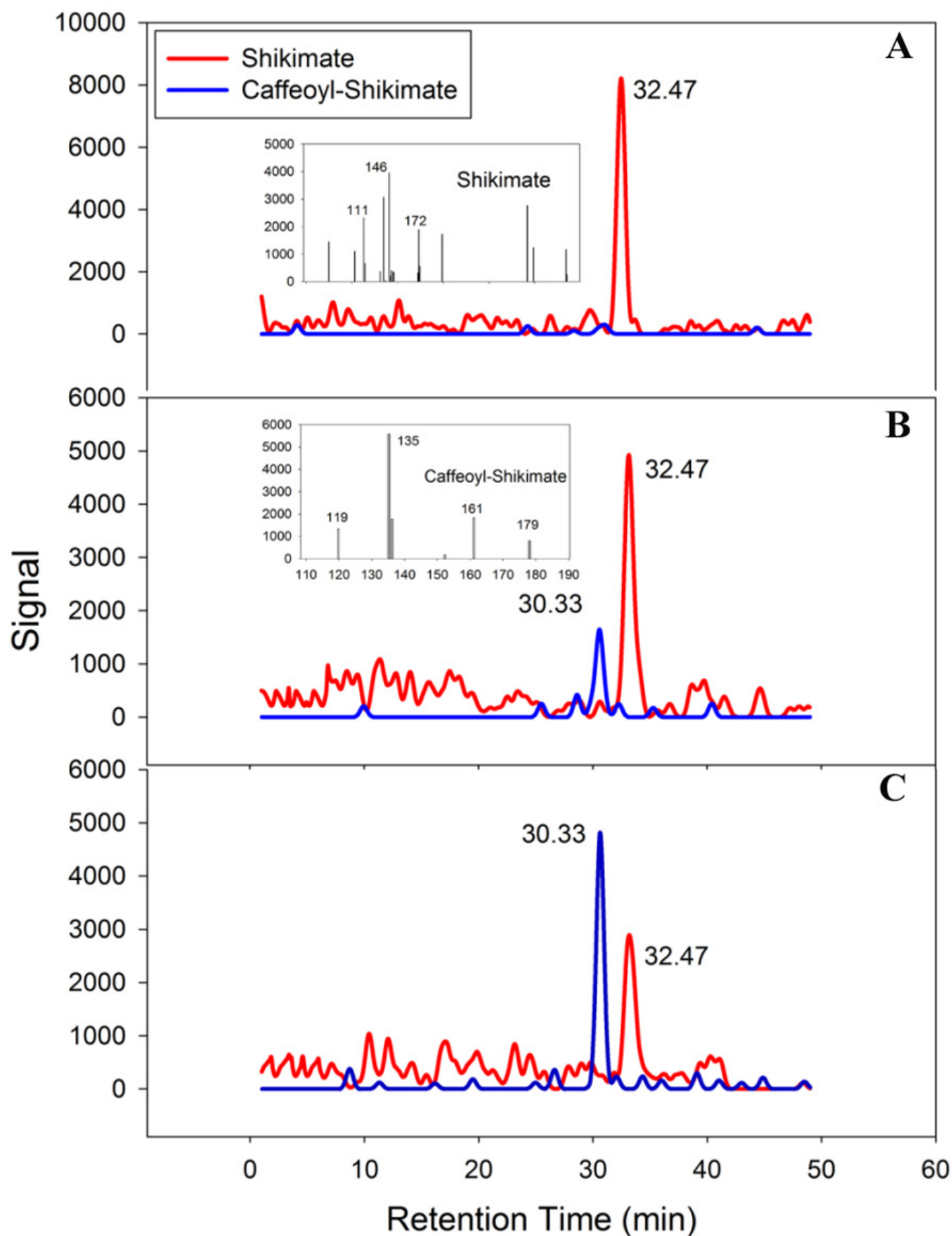
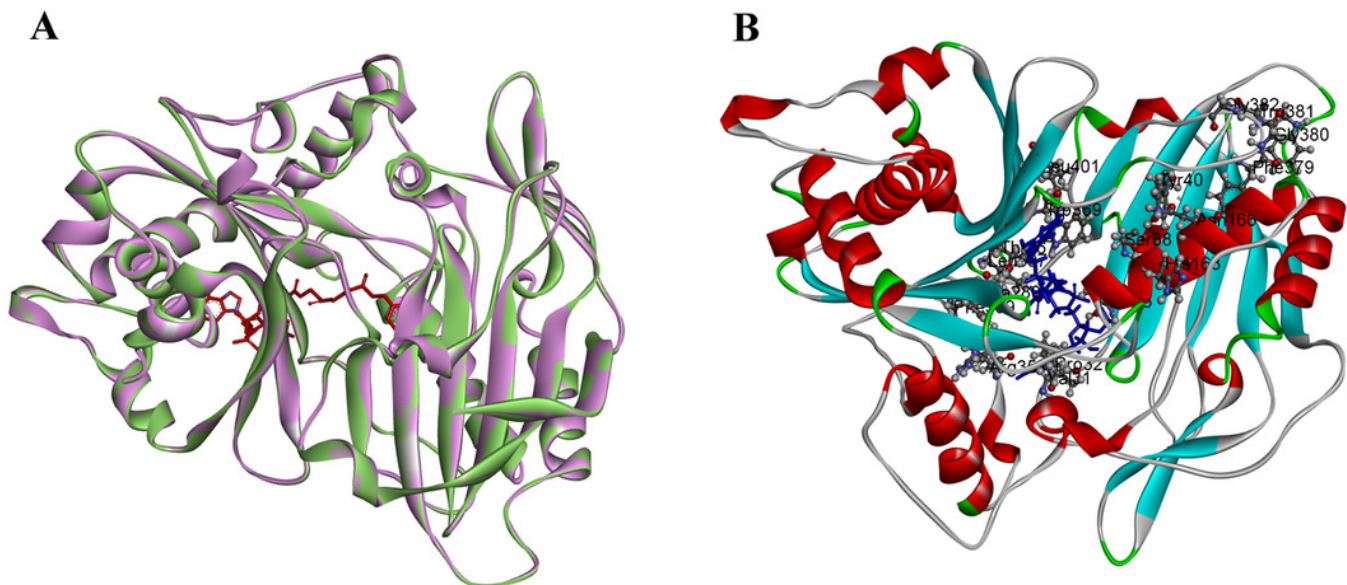



Figure 6

The structure of PtoHCT1 and docking with caffeoyl-CoA

(A) structure alignment of AtHCT (green) and PtoHCT (purple). (B) PtoHCT docked with caffeoyl-CoA. Blue ligand is caffeoyl-CoA and active sites are labeled.

Summary of active sites of HCTs

1   Table 1 Summary of active sites of HCTs

Position	Amino acid	Annotation	Reference
31	Val	Shikimate binding	Walker et al. 2013
32	Pro	Shikimate binding	Walker et al. 2013
298	Ala	Shikimate binding	Walker et al. 2013
318	Ile	Shikimate binding	Walker et al. 2013
376	Phe	Shikimate binding	Walker et al. 2013
414	Leu	Shikimate binding	Walker et al. 2013,Lallemand et al. 2012
418	Leu	Shikimate binding	Walker et al. 2013
38	Ser	carbonyl group of p- coumaroyl moiety	Walker et al. 2013,Eudes et al. 2016
40	Tyr	carbonyl group of p- coumaroyl moiety	Walker et al. 2013,Lallemand et al. 2012,Eudes et al. 2016
384	Trp	carbonyl group of p- coumaroyl moiety	Walker et al. 2013
163	His	carbonyl group of shikimate moiety and Catalysis	Walker et al. 2013,Lallemand et al. 2012,Eudes et al. 2016
369	Arg	carbonyl group of shikimate moiety	Walker et al. 2013
382	Thr	carbonyl group of shikimate moiety	Walker et al. 2013,Eudes et al. 2016
36	Thr	Catalysis	Walker et al. 2013

2   Note: all positions correspond to SbHCT