Supporting Information:

Evolution of *p*-coumaroylated lignin in eudicots provides new tools for cell wall engineering

Yaseen Mottiar, Rebecca A. Smith, Steven D. Karlen, John Ralph, and Shawn D. Mansfield

Article Accepted: September 21, 2022

Table S1. RNA-seq analysis of putative homologues of monocot PMTs expressed in kenaf stems used to select candidate gene, shown in blue.

Contig ID:	RPKM:	Total reads:	Closest Arabidopsis Homologue:	Corresponding Annotation:	Additional Description:
contig_14943	4.27	2	AT1G03010	None	Phototropic-responsive NPH3 family protein
contig_13644	0	0	AT1G03390 (weak hit)	None	HXXXD-type acyl-transferase family protein
contig_27028	3.46	2	AT1G25580 (weak hit)	Arabidopsis NAC domain containing protein 8, SUPPRESSOR OF GAMMA RADIATION 1 (ANAC008, SOG1)	Encodes suppressor of gamma response 1 (SOG1), a putative transcription factor governing multiple responses to DNA damage.
contig_8764	12.23	8	AT1G27620	None	HXXXD-type acyl-transferase family protein
contig_18803	7.1	3	AT1G27620		
	19.33	11			
contig_1632	14.09	12	AT1G28680	COUMARIN SYNTHASE (COSY)	Catalyses trans-cis isomerization and lactonization in the biosynthesis of coumarins in roots.
contig_16003	21.12	9	AT1G53720 (weak hit)	CYCLOPHILIN 59, cyclophilin 59 (CYP59, ATCYP59)	Encodes a cyclophilin, member of a family modular proteins consisting of a peptidyl-prolyl cis– trans isomerase (PPIase) domain, followed by an RNA recognition motif (RRM), and a C-terminal domain enriched in charged amino acids.
contig_18932	4.73	2	AT1G65450	GLAUCE (GLC)	Contains dual transcription units and alternative splicing that could rescue the sterility defect of glc mutants. Shares homology to BAHD (for BEAT, AHCT, HCBT, and DAT) acyl-transferases. Functions in double fertilization.
contig_26826	9.81	5	AT1G69990 (weak hit)	BAK1-interacting receptor-like kinase 4 (BIR4)	Leucine-rich repeat protein kinase family protein
contig_339	513.13	800	AT2G25150 (weak hit)	None	HXXXD-type acyl-transferase family protein
contig_14098	35.58	35	AT2G37110	None	PLAC8 family protein
contig_25222	6.49	3	AT2G40230	None	HXXXD-type acyl-transferase family protein
contig_9677	79.49	62	AT3G03480 (weak hit)	acetyl CoA:(Z)-3-hexen-1-ol acetyltransferase (CHAT)	acetyl CoA:(Z)-3-hexen-1-ol acetyltransferase
contig_20086	6.78	3	AT3G23840	CER26-LIKE (CER26-LIKE, AtPNP-R1)	HXXXD-type acyl-transferase family protein
contig_6580	11.89	8	AT3G51150	None	ATP binding microtubule motor family protein
contig_17952	4.08	4	AT3G62160	None	HXXXD-type acyl-transferase family protein
contig_16976	24.54	35	AT4G08150	KNOTTED-like from Arabidopsis thaliana, BREVIPEDICELLUS 1 (KNAT1, BP1)	A member of class I knotted1-like homeobox gene family (together with KNAT2). Similar to the knotted1 (kn1) homeobox gene of maize.
contig_6826	16.04	12	AT4G15390	None	HXXXD-type acyl-transferase family protein
contig_17786	7.1	7	AT4G22140	EARLY BOLTING IN SHORT DAYS (EBS)	Encoding a chromatin remodelling factor that regulates flowering time.

Note: This table continues on the following page.

Note: Table continued from the preceding page.

Contig ID:	RPKM:	Total reads:	Closest Arabidopsis Homologue:	Corresponding Annotation:	Additional Description:
contig_4757	100.62	79	AT4G26000 (weak hit)	PEPPER (PEP)	Encodes a novel Arabidopsis gene encoding a polypeptide with K-homology (KH) RNA-binding modules, which acts on vegetative growth and pistil development. Genetic studies suggest that PEP interacts with element(s) of the CLAVATA signaling pathway.
contig_22124	6.05	6	AT5G25060 & AT5G10800	reduced red-light responses in cry1cry2 background 1 (RRC1)	RNA recognition motif (RRM)-containing protein
contig_11413	11.48	3	AT5G13230 (weak hit)	None	Tetratricopeptide repeat (TPR)-like superfamily protein
contig_2673	**30.17**	**24**	**AT5G17540**	**None**	**HXXXD-type acyl-transferase family protein**
contig_10163	**71.97**	**29**	**AT5G17540**		
contig_10499	**30.03**	**27**	**AT5G17540**		
contig_10500	**19.89**	**11**	**AT5G17540**		
contig_10501	**10.66**	**4**	**AT5G17540**		
	162.72	**95**			
contig_4167	8.45	5	AT5G22250	CCR4- associated factor 1b (AtCAF1b, CAF1b)	Encodes one of the homologs of the yeast CCR4-associated factor 1: AT3G44260 (CAF1a), AT5G22250 (CAF1b). Has mRNA deadenylation activity. Also plays a role in plant defense responses.
contig_22956	5.52	2	AT5G25300 (weak hit)	F-Box/DUF295 Brassiceae-specific 36 (AtFDB36)	F-box protein
contig_4656	14.94	17	AT5G37930	None	Protein with RING/U-box and TRAF-like domain
contig_3391	11.8	10	AT5G39670	CALMODULIN-LIKE 46 (CML46)	Calmodulin like protein involved in negative regulation of pattern triggered immunity.
contig_21422	11.42	4	AT5G41040	hydroxycinnamoyl- CoA:ω-hydroxyacid O-hydroxycinnamoyltransferase, aliphatic suberin feruloyl-transferase, REDUCED LEVELS OF WALL-BOUND PHENOLICS 1 (HHT, ASFT, RWP1)	Encodes a feruloyl-CoA transferase required for suberin synthesis. Has feruloyl-CoA-dependent feruloyl transferase activity towards substrates with a primary alcohol.
contig_1217	387.6	515	AT5G48930	hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT)	At5g48930 has been shown to encode for the hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase (HCT) both synthesizing and catabolizing the hydroxycinnamoylesters (coumaroyl/caffeoyl shikimate and quinate) involved in the phenylpropanoid pathway.
contig_1568	45.63	62	AT5G48930		
contig_2674	153.56	188	AT5G48930		
contig_8908	60.43	39	AT5G48930		
contig_18207	6.07	3	AT5G48930		
	653.29	807			

Table S2. Primers and reaction conditions for construct assembly, transformant screening, and qRT-PCR.

Description:	**Primer pair (5' → 3'):**	**PCR reaction conditions:**
Amplification of the synthetic *Hc*PMT coding sequence (1335 bp)	Forward primer: ATG GCA CTG CTA CGA CCG Reverse primer: CTA AGC ACC AGT TGC TTC ATC	2.5 µL ..... 10× buffer 0.5 µL ..... 50 mM $MgSO_4$ 0.5 µL ..... 10 µM forward primer 0.5 µL ..... 10 µM reverse primer 0.5 µL ..... 10 mM dNTPs mixture 0.5 µL ..... template DNA* 0.2 µL ..... *Pfx* DNA polymerase Up to 25 µL ..... deionised water With thermal cycling as follows: 94°C for 30 s, 34 cycles of (94°C for 15 s, 60°C for 30 s, 68°C for 90 s), and then 68°C for 10 min. * Note: template DNA was the synthetic *Hc*PMT coding sequence.
Addition of the *attB* adaptor sequences to enable Gateway cloning	Forward primer: GGG GAC AAG TTT GTA CAA AAA AGC AGG CTA TGG CAC TGC TAC GAC CG Reverse primer: GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TAA GCA CCA GTT GCT TCA TC	2.5 µL ..... 10× buffer 0.5 µL ..... 50 mM $MgSO_4$ 0.5 µL ..... 10 µM forward primer 0.5 µL ..... 10 µM reverse primer 0.5 µL ..... 10 mM dNTPs mixture 0.5 µL ..... template DNA* 0.2 µL ..... *Pfx* DNA polymerase Up to 25 µL ..... deionised water With thermal cycling as follows: 94°C for 30 s, 34 cycles of (94°C for 15 s, 60°C for 30 s, 68°C for 90 s), and then 68°C for 10 min. * Note: template DNA was a plasmid containing the *Hc*PMT coding sequence.
PCR-screening to identify positive transformants	Forward primer: CTG ACC GCA TGT GTT TGG Reverse primer: CTA AGC ACC AGT TGC TTC ATC Amplicon size: 543 bp	2.5 µL ..... 10× buffer 0.5 µL ..... 50 mM $MgSO_4$ 0.5 µL ..... 10 µM forward primer 0.5 µL ..... 10 µM reverse primer 0.5 µL ..... 10 mM dNTPs mixture 0.5 µL ..... template DNA 0.5 µL ..... *Taq* DNA polymerase Up to 25 µL ..... deionised water With thermal cycling as follows: 94°C for 3 min, 34 cycles of (94°C for 30 s, 55°C for 30 s, 72°C for 30 s), and then 72°C for 5 min. * Note: template DNA was genomic DNA preparations from young poplar leaves.

Note: This table continues on the following page.

Note: Table continued from the preceding page.

Description:	Primer pair (5' → 3'):	PCR reaction conditions:
Quantification of expression levels using RT-qPCR	For *Hc*PMT transgene, Forward primer: GCC TGC AAG CGT TAT GGA AA Reverse primer: GCT TCA TCA TCT GCC AGG GT Amplicon size: 63 bp For elongation factor 1β (reference), Forward primer: GGC ATT AAG TTT TGT CGG TCT G Reverse primer: GCG GTT CAT CAT TTC ATC TGG Amplicon size: 97 bp For actin (reference), Forward primer: ACC AGT GTG TCT TGG TCT ACC C Reverse primer: CGA TGC CGA GGA TAT TCA AC Amplicon size: 127 bp	5 µL ....BrightGreen qPCR master mix 0.3 µL .....10 µM forward primer 0.3 µL .....10 µM reverse primer 1 µL .....template cDNA Up to 10 µL.....deionised water With qPCR parameters as follows: 95°C for 30 s, 39 cycles of (95°C for 5 s, 60°C for 15 s, plate read for fluorescence), and then followed by a melt-curve analysis as follows: 95°C for 10 s, ramp from 55°C to 95°C at 0.5°C per increment with hold times of 5 s and a plate read for fluorescence. * Note: template DNA was cDNA preparations synthesised with RNA extracted from developing poplar xylem.

Table S3. Composition of structural polysaccharides (neutral sugars) in transgenic poplars expressing *Hc*PMT. The values shown in bold are significantly different from the wild-type (WT) control (one-way ANOVA with post-hoc Dunnett's test, n=5 for each line, p-value < 0.05). Mean values are provided along with standard deviation.

Sugar (% w/w)	Line 1	Line 2	Line 3	Line 4	Line 5	WT
Glucose	**41.25 ± 3.13**	44.30 ± 2.97	44.30 ± 0.90	44.62 ± 0.93	46.60 ± 3.41	45.21 ± 0.79
Xylose	15.76 ± 3.53	17.78 ± 0.65	18.26 ± 0.30	18.18 ± 0.22	18.23 ± 0.34	17.64 ± 0.30
Mannose	1.45 ± 0.19	1.67 ± 0.16	1.60 ± 0.09	1.58 ± 0.06	**1.80 ± 0.08**	1.57 ± 0.10
Galactose	0.79 ± 0.05	0.74 ± 0.02	0.75 ± 0.05	0.74 ± 0.07	0.70 ± 0.09	0.77 ± 0.08
Rhamnose	0.44 ± 0.02	0.44 ± 0.01	**0.46 ± 0.01**	0.43 ± 0.02	0.43 ± 0.01	0.43 ± 0.01
Arabinose	0.38 ± 0.03	0.33 ± 0.01	0.36 ± 0.04	0.35 ± 0.02	**0.27 ± 0.03**	0.34 ± 0.02

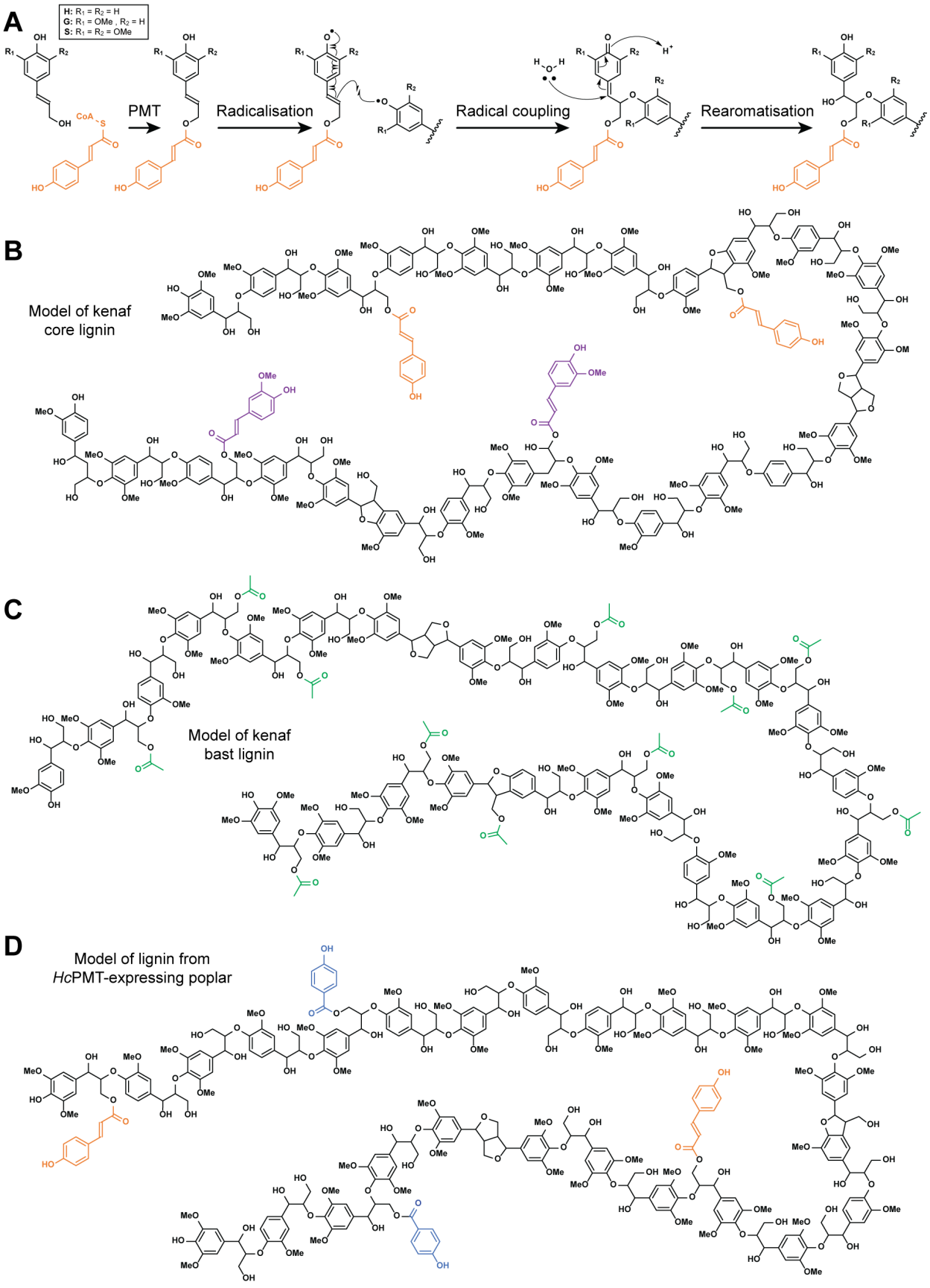
A
H: R1 = R2 = H
G: R1 = OMe , R2 = H
S: R1 = R2 = OMe
PMT
Radicalisation
Radical coupling
Rearomatisation
B
Model of kenaf
core lignin
C
Model of kenaf
bast lignin
D
Model of lignin from
HcPMT-expressing poplar

Figure S1. Incorporation of ester-linked pendent groups into lignin. **A.** Conjugates formed by *p*-coumaroyl-CoA:monolignol transferase (PMT) enzymes can participate in radical coupling and lignification, as shown. **B.** A hypothetical model of lignin from kenaf core tissues depicting *p*-coumarate groups in orange and ferulate groups in magenta. **C.** A hypothetical model of lignin from kenaf bast fibres depicting acetate groups in green. **D.** A hypothetical model of lignin from transgenic poplar expressing *Hc*PMT depicting novel *p*-coumarate groups in orange, and native *p*-hydroxybenzoate groups in blue.

>*Hc*PMT_cds_reported_in_Zhang_et_al_2020
ATGGCACTGCTACGACCCGCTTCATTAGTGTTCACTGTACGGAGGCATGATCCGGAGCTTGTCGTCCCGTCTAAACC
AACCCCTCACGAGTGTAAGACGCTGTCGGACATCGACGACCAAGACGGTCATCGTTTTCAAATCCGAGGGCTCCATG
TTTACCGGTGCAACGCTAGCATGCAAGGAAAGGATCCCGTTAGGGTTATAAGAGAAGCACTTGCTAAAGCCCTAGTG
TTTTATTATCCATTTGCTGGTAGAATAAAGGAAGGGCCGAACCGGAAGCTTATGGTGGATTGTACTGGGGAAGGTGT
GTTGTTTATCGAGGCCGATGCTGATGTTATGCTCGAGGAATTTGGTGGTTCACTTCATCCTCCATTTCCATGTTTCA
AGGATCTTCTTTGTGAGCCCACTGGCTCAAATGATTTGTTAAACTCTCCGGTGCTACAAATTCAGGTGACACGCTTG
AAATGTGGCGGTTTCATCTTTGCGCACCGATTCAACCACACCATGAGCGACGCTGTCGGGCTGATTCAATTCATGTC
TGCGATGGGCGAAATAGCACGAGGTGCAGTGGCTCCCTCAATCCCACCCGTTTGGGAAAGACATCTTTTGAATGCTC
GAGACCCGCCGCTCATAACATGCGAGCACCACGAGTACGACCATGCCACGGCCACCAACGGTACAATCATGCCCACA
GACAACTTGGTTCACCACTCTTTCTTCTTCGGGCCAACTCAAATTTCAGCTCTTAAAAGACTCATCTCCGATAATGT
CAGTTGTTCAACGTTCGATATCTTAACCGCATGCGTGTGGCGCTGTCGTACGATAGCCATGAAACTCGGCCCCGACG
AAGACGTTCGGCTCATATGCATAGTCAATGCAAGGTACAAATTCAACCCTCCTTTGCCATTAGGGTATTACGGGAAT
GCACTCGGATACCCCGCGGCTCTAACAACCGCAGGCGAACTTAGCAAAAAACCATTAGAATACGCAGTAAAGCTAGT
GAAGGAGGCCAAGGCAAAGGCTACAGACGAGTACATGAAATCGACGGCGGATTTGATGGTGAGTCGAGGACGGCCAA
ACGTTAACACGGTTCGATCTTTTCTGGTGTCGGATTTGAGTCGGGCCAGGTTCAGGGAGGTTGACTTTGGGTGGGGT
AAGGCAGAGTTTGGTGGACCGTCGAATGGTACAGAGATTATAAGCTTTTATATACCTTCAAAGAATAAGGAAGGAAA
GGAAGGGATTGCAGTGCCGGTTTGCTTGCCTGCATCTGTTATGGAGAGCTTTGTTAAAGAGATTAATTCAACGCTGG
CAGACGATGAAGCAACTGGTGCTTAG

>*Hc*PMT_amino_acid_sequence
MALLRPASLVFTVRRHDPELVVPSKPTPHECKTLSDIDDQDGHRFQIRGLHVYRCNASMQGKDPVRVIREALAKALV
FYYPFAGRIKEGPNRKLMVDCTGEGVLFIEADADVMLEEFGGSLHPPFPCFKDLLCEPTGSNDLLNSPVLQIQVTRL
KCGGFIFAHRFNHTMSDAVGLIQFMSAMGEIARGAVAPSIPPVWERHLLNARDPPLITCEHHEYDHATATNGTIMPT
DNLVHHSFFFGPTQISALKRLISDNVSCSTFDILTACVWRCRTIAMKLGPDEDVRLICIVNARYKFNPPLPLGYYGN
ALGYPAALTTAGELSKKPLEYAVKLVKEAKAKATDEYMKSTADLMVSRGRPNVNTVRSFLVSDLSRARFREVDFGWG
KAEFGGPSNGTEIISFYIPSKNKEGKEGIAVPVCLPASVMESFVKEINSTLADDEATGA

Figure S2. Nucleotide and amino acid sequences of *Hc*PMT.

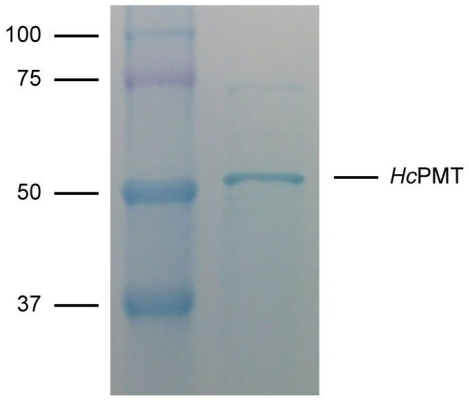

Figure S3. Coomassie-stained SDS-PAGE gel. Recombinant *Hc*PMT enzyme was purified from *E. coli* by immobilised metal affinity chromatography. The sizes of molecular weight standards (Precision-Plus Dual Colour Standards, Bio-Rad Labs) are labelled on the left in kDa.

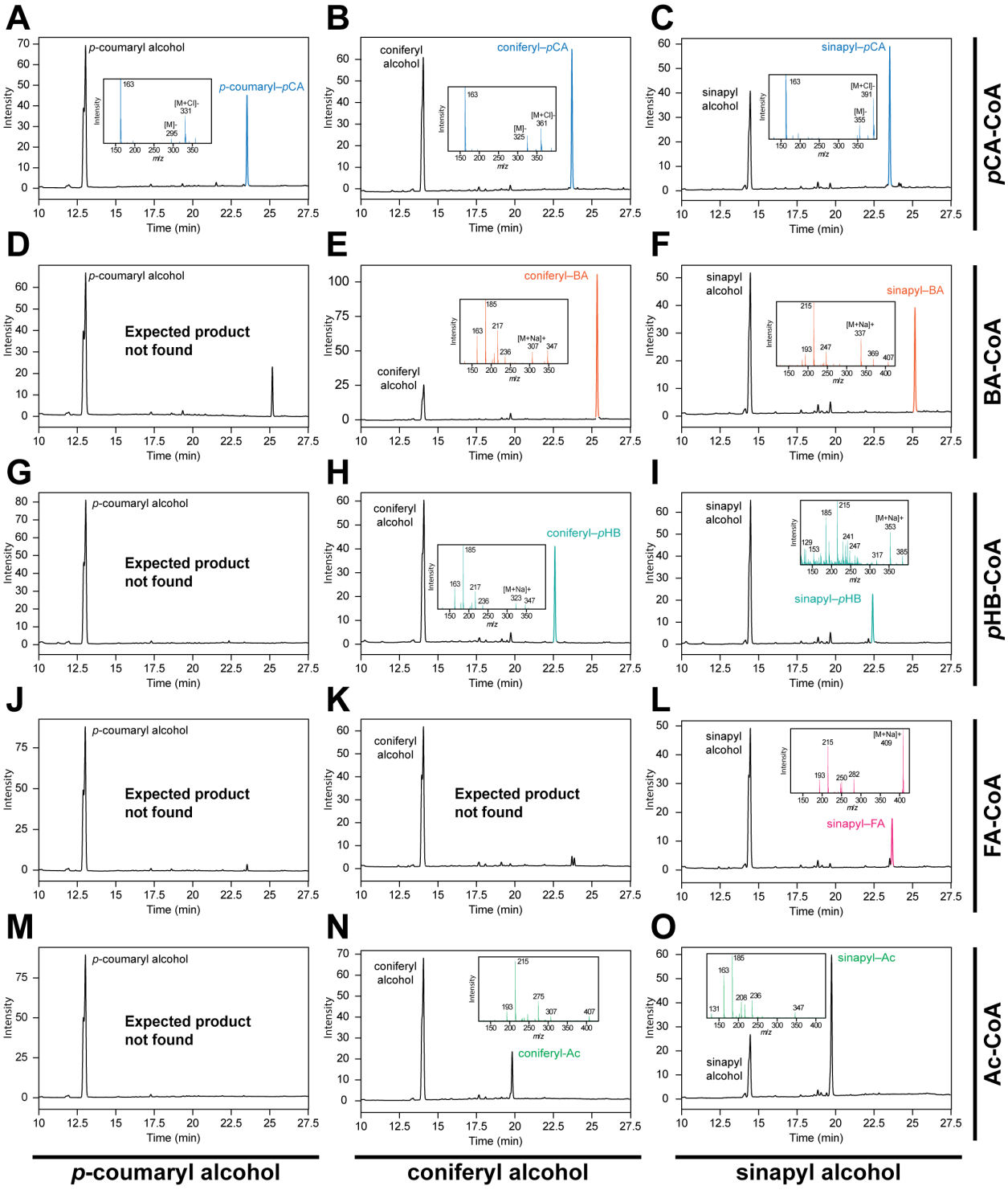
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
p-coumaryl alcohol
coniferyl alcohol
sinapyl alcohol
p-coumaryl–pCA
coniferyl–pCA
sinapyl–pCA
coniferyl–BA
sinapyl–BA
coniferyl–pHB
sinapyl–pHB
sinapyl–FA
coniferyl-Ac
sinapyl–Ac
Expected product not found
Intensity
Time (min)
m/z
pCA-CoA
BA-CoA
pHB-CoA
FA-CoA
Ac-CoA
p-coumaryl alcohol
coniferyl alcohol
sinapyl alcohol

Figure S4. Pairwise assays to screen *Hc*PMT enzyme activity. **A-C.** Assays of *p*CA-CoA with the three monolignols produced *p*-coumaryl–*p*CA, coniferyl–*p*CA, and sinapyl–*p*CA. **D-F.** Assays with BA-CoA yielded coniferyl–BA and sinapyl–BA. **G-I.** Assays with *p*HB-CoA yielded coniferyl–*p*HB and sinapyl–*p*HB. **J-L.** Assays with FA-CoA yielded only sinapyl–FA. **M-O.** Assays with Ac-CoA yielded coniferyl–Ac and sinapyl–Ac. Chromatographic traces show the absorbance at 273 nm. Insets show mass fragmentation patterns for monolignol conjugates, which were validated using authentic standards. All reactions were performed at room temperature for 60 min with 1 µg purified enzyme, 1 mM dithiothreitol, 1 mM of each CoA thioester, and 1 mM of each monolignol in 50 mM phosphate buffer (pH 6).

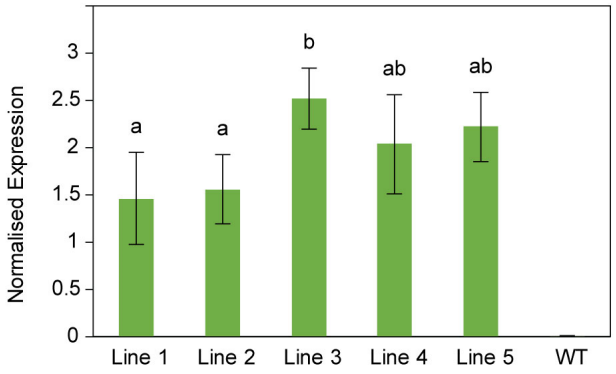

Figure S5. Transgene expression in transgenic poplar expressing *Hc*PMT. Relative expression of *Hc*PMT, shown in green, was measured by qRT-PCR. No expression was detected for the wild-type (WT) control. Expression was normalised relative to the poplar actin and elongation factor 1β sequences. Five biological replicates were analysed for each line using technical triplicates. The letters above each bar indicate statistically significant differences between the transgenic lines (one-way ANOVA with Tukey-Kramer test, $\alpha < 0.05$). Error bars represent standard deviation.

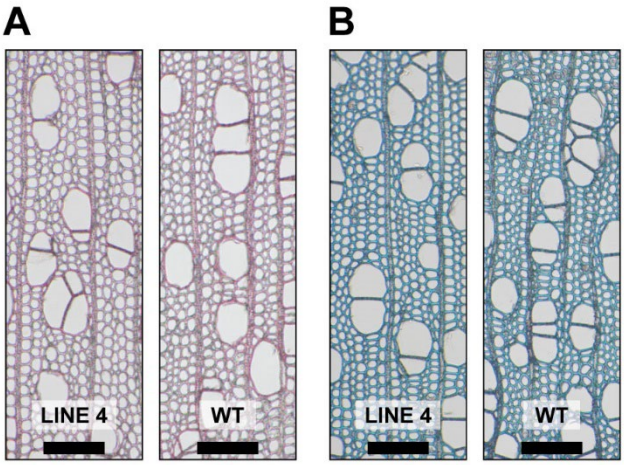

Figure S6. Light microscopy of representative polar stems. **A**. Transverse stem cross-sections of transgenic poplar expressing *Hc*PMT (line 4) and the wild-type (WT) control, as labelled, stained with phloroglucinol-HCl. **B**. The same, except stained with toluidine blue. The black scale bars represent 100 μm.

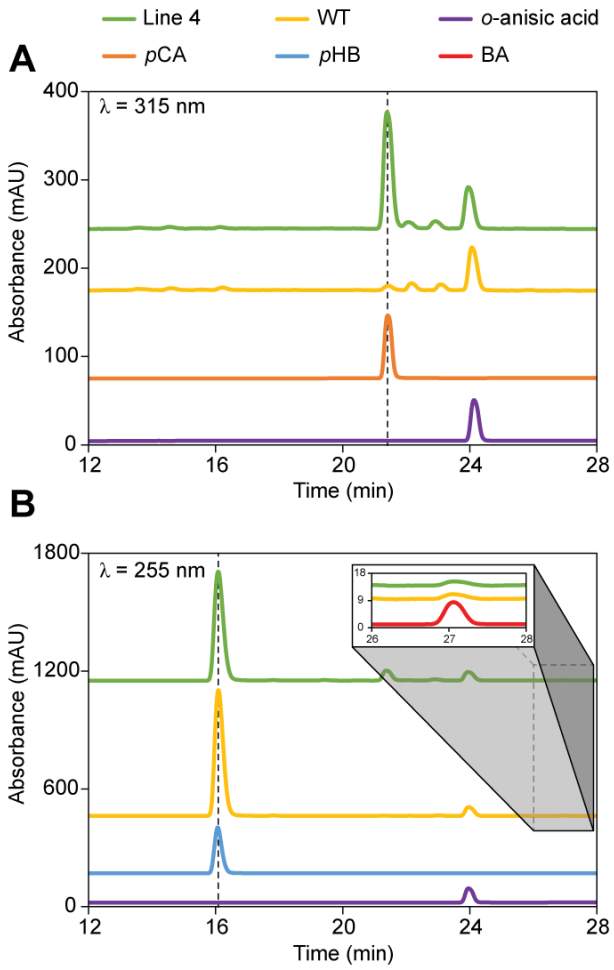

Figure S7. Analysis of cell-wall-bound phenolics in transgenic poplar expressing *Hc*PMT by high-performance liquid chromatography. **A.** Alkaline hydrolysates of extractive-free wood released substantial amounts of *p*-coumarate (*p*CA) from line 4 (green) and only trace amounts from the wild-type (WT) control (yellow). Detection was performed at 315 nm, the λ_{max} for *p*CA. **B.** The same as in A, but with detection at 255 nm, the λ_{max} for *p*-hydroxybenzoate (*p*HB), showing the slight decrease in *p*HB released from line 4 relative to the WT control. Vertical dashed lines mark the peaks corresponding to *p*CA, in **A**, and *p*HB, in **B**. The shaded inset in **B** provides a zoomed-in view of 26–28 min and shows that only trace amounts of benzoate (BA) were detected. Chromatograms are also presented for authentic *p*CA (orange), *p*HB (blue), BA (red), as well as *o*-anisic acid (purple) which was used as an internal standard.

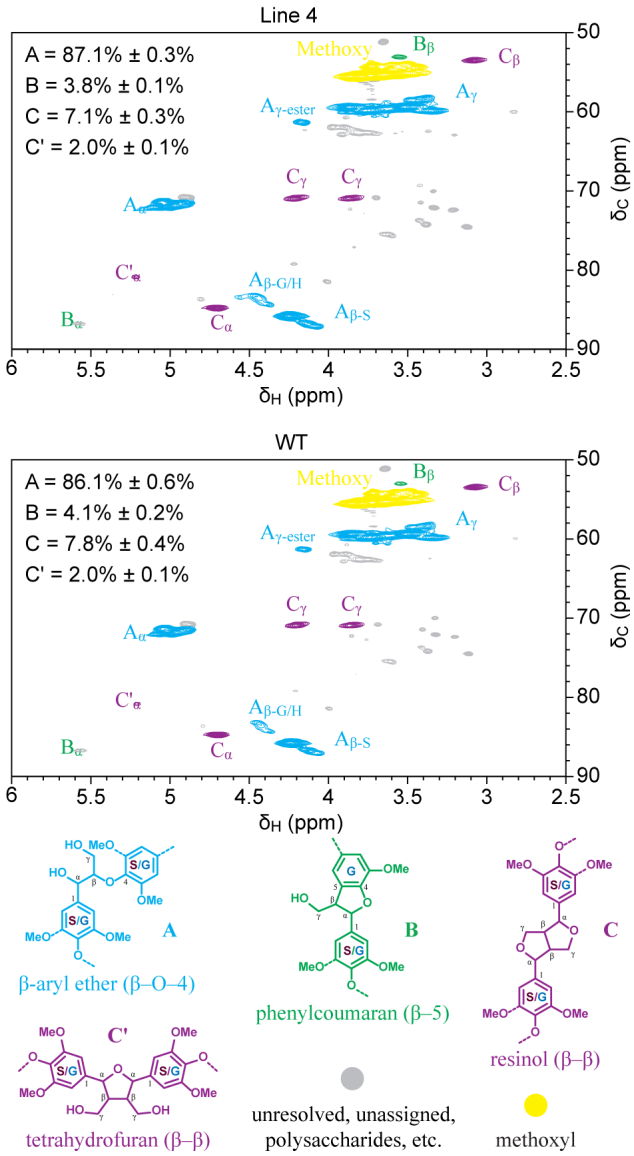

Figure S8. Additional NMR spectra of transgenic poplar expressing *Hc*PMT. Two-dimensional ^{1}H–^{13}C HSQC NMR spectra for enzyme-lignin samples of line 4 and the wild-type (WT) control, as labelled, showing the aliphatics region. The colour-coding and peak annotations are elaborated with the structures shown below for β-aryl ether (A, cyan), phenylcoumaran (B, green), resinol (C, magenta), and tetrahydrofuran units (C', magenta). Proportions for integrated peak volumes are provided as the mean ± standard error for three biological replicates.

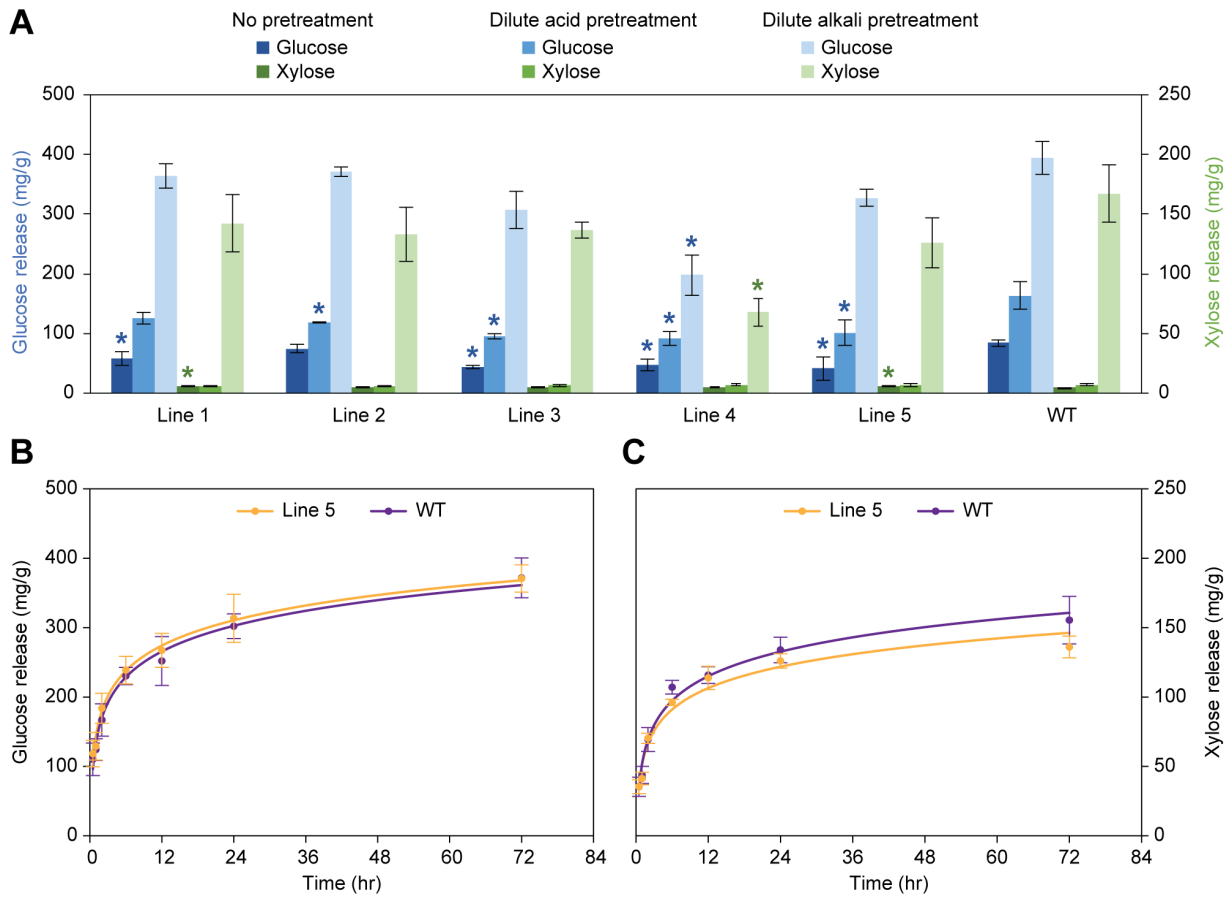

Figure S9. Saccharification potential of transgenic poplar expressing *Hc*PMT. **A.** The release of glucose (shades of blue) and xylose (shades of green) is shown following various pretreatments (dilute alkali, dilute acid, and no pretreatment, as labelled) and a 72-h enzymatic hydrolysis. Those values marked with an asterisk are significantly different from the wild-type (WT) control (one-way ANOVA with Dunnett's test, n=3 for each line, p-value < 0.05). Error bars represent standard deviation. **B.** Enzymatic hydrolysis time course showing the release of glucose for line 5 (orange) and the WT control (magenta) following dilute alkali pretreatment. **C.** The same as **B,** but showing xylose release. For **A**–**C**, the error bars show standard deviation for three biological replicates, each measured in triplicate.

SI References

Zhang L, Xu Y, Zhang X, Ma X, Zhang L, Liao Z, Zhang Q, Wan X, Cheng Y, Zhang J, Li D, Zhang L, Xu J, Tao A, Lin L, Fang P, Chen S, Qi R, Xu X, Qi J, Ming R. 2020. The genome of kenaf (*Hibiscus cannabinus* L.) provides insights into bast fibre and leaf shape biogenesis. *Plant Biotech. J.* 18(8): 1796–1809.