Shi Hu (Orcid ID: 0000-0002-9067-0766)
Takata Naoki (Orcid ID: 0000-0002-8479-8165)
Kajita Shinya (Orcid ID: 0000-0001-7187-6059)

Rerouting of the lignin biosynthetic pathway by inhibition of cytosolic shikimate recycling in transgenic hybrid aspen

Authors:

Shi Hu[1], Naofumi Kamimura[2], Shingo Sakamoto[3,4], Soichiro Nagano[5], Naoki Takata[6], Sarah Liu[7], Geert Goeminne[8,9], Ruben Vanholme[8,9], Mikiko Uesugi[1], Masanobu Yamamoto[1], Shojiro Hishiyama[10], Hoon Kim[7], Wout Boerjan[8,9], John Ralph[7], Eiji Masai[2], Nobutaka Mitsuda[3,4], Shinya Kajita[1,*]

[1] Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan

[2] Department of Bioengineering, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan

[3] Plant Gene Regulation Research Group, Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan

[4] Smart CO2 Utilization Research Team, Global Zero Emission Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan

[5] Forest Tree Breeding Center, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Hitachi, Ibaraki 319-1301, Japan

[6] Forest Bio-Research Center, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Hitachi, Ibaraki 319-1301, Japan

[7] Department of Biochemistry, and U.S. Department of Energy Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, University of Wisconsin, Madison, WI, 53726 USA

[8] Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium

[9] VIB Center for Plant Systems Biology, 9052 Ghent, Belgium

[10] Department of Forest Resource Chemistry, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Tsukuba, Ibaraki 305-8687, Japan

Running title: Role of shikimate recycling in lignin biosynthesis

Keywords: heterologous expression, lignocellulose, metabolome, phenolic metabolites, RNA-seq, shikimate pathway, saccharification efficiency

[*] **Corresponding author**
Shinya Kajita
Graduate School of Bio-Applications and Systems Engineering
Tokyo University of Agriculture and Technology
2-24-16 Nakacho, Koganei, Tokyo 184-8588, Japan

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.15674

FAX, +81 42 388 7391. E-mail, kajita@cc.tuat.ac.jp

Summary

Lignin is a phenolic polymer deposited in the plant cell wall and is mainly polymerized from three canonical monomers (monolignols), i.e., *p*-coumaryl, coniferyl, and sinapyl alcohols. After polymerization, these alcohols form different lignin substructures. In dicotyledons, monolignols are biosynthesized from phenylalanine, an aromatic amino acid. Shikimate acts at two positions in the route to the lignin building blocks. It is part of the shikimate pathway which provides the precursor for the biosynthesis of phenylalanine, and is involved in the transesterification of *p*-coumaroyl-CoA to *p*-coumaroyl shikimate, one of key steps in the biosynthesis of coniferyl and sinapyl alcohols. The shikimate residue in *p*-coumaroyl shikimate is released in later steps, and the resulting shikimate becomes available again for the biosynthesis of new *p*-coumaroyl shikimate molecules. In this study, we inhibited cytosolic shikimate recycling in transgenic hybrid aspen by accelerated phosphorylation of shikimate in the cytosol through expression of a bacterial shikimate kinase. This expression elicited an increase in *p*-hydroxyphenyl units of lignin and, by contrast, a decrease in guaiacyl and syringyl units. Transgenic plants with high shikimate kinase activity produced a lignin content comparable to that in wild-type plants and had an increased processability via enzymatic saccharification. Although expression of many genes was altered in the transgenic plants, elevated shikimate kinase activity did not exert a significant effect on the expression of the majority of genes responsible for lignin biosynthesis. The present results indicate that cytosolic shikimate recycling is crucial to the monomeric composition of lignin rather than for lignin content.

Significance statement

To investigate the role of shikimate recycling in lignin biosynthesis, cytosolic shikimate availability was reduced by overexpression of a bacterial shikimate kinase in hybrid aspen. The transgenic trees showed an increase in *p*-hydroxyphenyl units of lignin without alteration of lignin level and the change in the structure led reduced biomass recalcitrance with higher delignification and improved cell wall digestibility.

Introduction

Shikimate is one of the metabolites in the shikimate pathway and contributes to the biosynthesis of aromatic amino acids such as phenylalanine and tyrosine. Over 30% of carbon fixed by plants passes through this pathway (Maeda and Dudareva, 2012). The pathway in plants comprises seven reaction steps that occur in the chloroplast. Shikimate also plays an important role as a key metabolite in the lignin biosynthetic pathway in the cytosol (Figure 1) (Hoffmann *et al.*, 2003, Hoffmann *et al.*, 2004).

Lignin is one of the main components of the plant cell wall, and it is a phenolic polymer with complex composition and structure (Boerjan *et al.*, 2003, Barros *et al.*, 2015). Deposition of lignin in the cell wall contributes to plant growth, tissue and organ development, lodging resistance, and resilience to various biotic and abiotic stresses (Liu *et al.*, 2018). This polymer is composed of distinct major structural units, i.e., *p*-hydroxyphenyl (H), guaiacyl (G), and

syringyl (S) units, that are mainly derived from the monolignols *p*-coumaryl, coniferyl, and sinapyl alcohols, respectively (Freudenberg and Neish, 1968). As is the case with lignin content, the levels of these units vary among plant species and between organs, tissues, and cells of individual plants (Barros *et al.*, 2015, Vanholme *et al.*, 2019). Such differences significantly affect the chemical, biochemical, and physical characteristics of plant cell walls. Monolignols are phenylpropanoids produced via the monolignol biosynthetic pathway which comprises several steps catalyzed by different enzymes (Boerjan *et al.*, 2003). The abundance, the tissular and cellular distribution, and the kinetic properties of these enzymes all are important factors determining lignin content and its monomeric composition. With only few exceptions, downregulation and loss of function of the corresponding genes elicit changes in lignin content and in its monomeric composition.

Most enzymes in the general phenylpropanoid pathway are localized in the cytosol, and others such as cytochrome P450 enzymes that catalyze hydroxylation of aromatic rings in monolignol intermediates, are anchored to the cytosolic face of the endoplasmic reticulum (Takabe *et al.*, 2001, Sato *et al.*, 2004). Transesterification of *p*-coumaroyl-CoA with shikimate, the product of which is *p*-coumaroyl shikimate, allows 3'-hydroxylation of the ring and is therefore a crucial step for channeling the phenylpropanoid pathway towards production of coniferyl and sinapyl alcohols (Padmakshan *et al.*, 2021). The reaction is catalyzed by hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (HCT) which is presumably localized in the cytosol (Hoffmann *et al.*, 2003) and/or associated with the endoplasmic reticulum membranes (Bassard *et al.*, 2012). Shikimate is biosynthesized in the plastid, then likely transported to the cytosol through an uncharacterized membrane transporter, and is then utilized as a substrate for HCT. After 3'-hydroxylation of *p*-coumaroyl shikimate by *p*-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H), shikimate is released from the resulting caffeoyl shikimate either by HCT or caffeoyl shikimate esterase (CSE) and is assumed to be recycled in the cytosol for biosynthesis of further *p*-coumaroyl shikimate (Franke *et al.*, 2002b, Vanholme *et al.*, 2013b).

Downregulation of the *HCT* gene is predicted to reduce shikimate consumption and the flux of early metabolites to downstream precursors of monolignols in the cytosol. Such downregulation should also increase metabolite flux toward the H units of lignin. Indeed, an increase in H unit levels and lignin reduction have been achieved by *HCT*-suppression in alfalfa (*Medicago sativa* L.), loblolly pine (*Pinus taeda* L.), *Brachypodium distachyon* (Serrani-Yarce *et al.*, 2021) and poplar (*Populus alba* × *P. glandulosa*) hybrids (Shadle *et al.*, 2007, Wagner *et al.*, 2007, Zhou *et al.*, 2018). Moreover, fiber-specific knockout of *HCT* in *Arabidopsis thaliana* also induces lignin reduction with elevated H unit levels (Liang *et al.*, 2019). A natural poplar mutant partially deficient in *HCT1* also exhibited higher H unit proportions in lignin than the controls, whereas no significant reduction in lignin content was observed in the mutant (Vanholme *et al.*, 2013a).

C3'H downregulation also reduces the flux to G- and S-type precursors. As is the case with *HCT* suppression, downregulation of *C3'H* elicits simultaneous changes in H unit levels and

lignin content in transgenic plants. Up to 65% increase in H unit levels and a 30% decrease in lignin content have been reported in *C3'H*-downregulated alfalfa (Ralph *et al.*, 2006). A significant increase in H units (up to 20% of all released monomers) accompanied by marked reduction in lignin content (less than 50% compared to the wild type) was observed in *C3'H*-downregulated hybrid poplar (Coleman *et al.*, 2008). A *C3'H* mutant of *Arabidopsis* exhibited similar trends regarding H unit and lignin content, as observed in *C3'H*-downregulated alfalfa and hybrid poplar (Bonawitz *et al.*, 2014).

CSE suppression can be expected to inhibit both shikimate release from caffeoyl shikimate and the flux to downstream metabolites in the monolignol biosynthetic pathway. Loss of function or downregulation of the gene can cause significant lignin reduction in *Arabidopsis*, *Medicago truncatula*, and poplar (Vanholme *et al.*, 2013b, Ha *et al.*, 2016, Saleme *et al.*, 2017, de Vries *et al.*, 2021a). In contrast to lignin reduction, the relative level of H units in lignin increases substantially in *CSE* mutants of these plants, as is the case with *HCT* and *C3'H* downregulation.

Although the importance of enzymes that are involved in metabolism of shikimate and shikimate-containing precursors for lignin biosynthesis was clearly recognized in the studies cited above, the constitutive availability of shikimate in the cytosol to ensure homeostasis in lignin accumulation has received comparably little attention. To the best of our knowledge, no previous studies examined the role of cytosolic shikimate recycling in lignin biosynthesis. Eudes *et al.* (2015) expressed a bacterial gene for plastid-targeted 3-dehydroshikimate dehydratase (*QsuB*) that can convert 3-dehydroshikimate, the immediate precursor of shikimate in the shikimate pathway, to protocatechuate in transgenic *Arabidopsis*. Even though expression of *QsuB* resulted in increased H unit and reduced lignin levels in the transgenic plants, as observed in *HCT*-, *C3'H*-, and *CSE*-downregulated and/or knockout plants, it is unclear whether qualitative and quantitative changes in lignin observed in the plants expressing *QsuB* were caused by the reduction in shikimate in the cytosol, insufficient production of monolignol precursors such as cinnamate and *p*-coumarate that are downstream metabolites derived from shikimate, or both.

In order to better understand the role of cytosolic shikimate recycling with respect to determining lignin amount and composition, reducing shikimate availability in the cytosol is a promising approach. To this end, specific conversion of cytosolic shikimate to shikimate 3-phosphate (S3P) has been attempted here by overexpression of shikimate kinase (SK) in the cytosol. S3P is a direct product of shikimate in the endogenous shikimate pathway and, in plants, phosphorylation of shikimate occurs in the plastids (Maeda and Dudareva, 2012). Phosphatase activity against S3P was suggested to occur in the tonoplast or inside the vacuole, but not in the cytosol (Holländer-Czytko and Amrhein, 1983). These insights suggest that successful reduction of shikimate levels in the cytosol may be realized through artificial phosphorylation of shikimate.

In the current study, an SK originating from *Helicobacter pylori* (HpSK) with a low Michaelis constant (*K*m) against shikimate (60 μM; Cheng *et al.*, 2005), compared to other

known SKs, was selected for expression in a hybrid aspen (*Populus tremula* L. × *P. tremuloides* Michx., cv. T89) under the control of a xylem-preferential promoter to reduce the level of cytosolic shikimate. HpSK is one of the best characterized bacterial SKs because of its medical importance (Han *et al.*, 2007), and its *K*m against shikimate is significantly lower than those of HCTs in *Populus* and tobacco plants (895 and 750 μM against shikimate, respectively; Hoffmann *et al.*, 2003, Kim *et al.*, 2011). Our results suggest that cytosolic shikimate recycling is a crucial step for determining monomeric composition of lignin rather than for regulating lignin content in the cell wall.

Results

Overexpression of *HpSK* in hybrid aspen

A codon-optimized coding sequence of *HpSK* was expressed in hybrid aspen under the control of the xylem-preferential promoter of the cinnamate 4-hydroxylase (C4H) gene, originating from another hybrid aspen (Kawai *et al.*, 1996). We produced seven independent transgenic plant lines expressing the *HpSK* construct. Figure 2a shows transgenic plants grown for 15 weeks after transplantation to pots in a conditioned culture room. The morphology of five transgenic lines (lines #15, #22, #24, #26, and #27) was similar to that of the wild type; however, lines #13 and #21 had significantly reduced plant height (Table S1). Lines #13 and #21 also exhibited a smaller stem diameter compared to the wild type, and their stem dry weights were significantly lower (76% and 71%, respectively) than that of the wild type (Table S1). Reverse-transcription PCR (RT-PCR) indicated that *HpSK* expression in four transgenic lines (lines #13, #15, #21, and #22) was higher than that in the other two lines (lines #26 and #27; Figure 2b). Based on these results, the four lines with the highest *HpSK* expression (lines #13, #15, #21, and #22) and the wild type were selected for further analyses.

Sub-cellular localization of HpSK and levels of SK activity in the transgenic lines

To investigate whether expression from the HpSK constructs would result in cytoplasmic localization of the HpSK protein, the sub-cellular localization of HpSK was monitored in *Arabidopsis* mesophyll protoplasts that transiently expressed fusion proteins, wherein green fluorescent protein (GFP) was fused to the amino- and carboxyl-termini of HpSK (Yoshida *et al.*, 2013, Sakamoto *et al.*, 2016). In both cases, GFP fluorescence was localized predominantly in the cytoplasm as was the case when GFP was expressed alone (Figure S1).

SK activity was measured using crude extract prepared from stems of each of the four transgenic lines and of the wild type (Figure 3). These values are considered to represent the sum of the activity of both endogenous SK and HpSK. In wild-type plants, SK activity was detected, but the level was extremely low, as was the case in lines #15 and #22. In contrast, SK activity in lines #13 and #21 which showed reduced growth compared to the wild type, was significantly higher than that in the wild type and in the other two transgenic lines with normal growth.

We examined time-dependent phosphorylation of shikimate and simultaneous *p*-coumaroyl

shikimate formation using crude extracts prepared from the plants (Figure S2). In contrast to the weak shikimate consumption in the wild type, efficient conversion of shikimate to S3P occurred in extracts from lines #13 and #21, with most shikimate being consumed at 120 min in these *HpSK* expression lines (Figure S2b and c). Simultaneous formation of compound I (retention time 2.62 min) was detected in the same reaction mixtures at 20 min and its level had increased further at 120 min (Figure S2). Based on ultra-high-pressure liquid chromatography (UHPLC)–mass spectrometry (MS) analysis of the reaction product of a recombinant *Arabidopsis* HCT with shikimate and *p*-coumaroyl-CoA, compound I was identified as *p*-coumaroyl shikimate (Figure S3). The product levels of *p*-coumaroyl shikimate were higher in the wild type than in lines #13 and #21 at 120 min (Figure S2). Throughout the experiments, no peak assigned as direct reaction products formed from S3P and *p*-coumaroyl-CoA were detected (Figure S4). These results indicated that *HpSK* expression elicited stable shikimate phosphorylation and should thus reduce shikimate availability for the production of *p*-coumaroyl shikimate via HCT activity. Furthermore, after 20 min reaction time, when there was still excess of shikimate, the amount of *p*-coumaroyl shikimate produced was almost the same for each protein extract, suggesting that the abundance of HCT was similar in lines #13 and #21 as in the wild type.

Effect of SK expression on lignin content and composition, and cell-wall-bound phenolics

Table 1 shows the lignin content measured by two different methods and its monomeric composition monitored using the thioacidolysis procedure. Regardless of the SK activity of these lines, the lignin content measured by both methods did not differ significantly between the transgenic lines and the wild type. Furthermore, there was no significant difference in the molecular weight distribution of isolated lignins between the wild type and *HpSK* lines (Figure S5).

In contrast, the two transgenic lines (#13 and #21) with higher SK activity differed from the wild type in terms of monomeric composition of lignin (Table 1 and Figure S6). In general, the H unit levels in lignins of poplar and aspen are <1% (Yue *et al.*, 2012). As opposed to the wild type in which the H unit level was below the limit of detection via thioacidolysis, the levels in the two transgenic lines #13 and #21 were significantly elevated. Although the relative proportions of G and S monomers differed only slightly between the transgenic lines and the wild type, the detection level of both monomers decreased significantly in the two lines (Table 1 and Figure S6). An increase in degradative products with H units and a decrease in the levels of G and S monomers were also detected in lines #13 and #21 by pyrolysis-gas chromatography–mass spectrometry (Py-GC–MS) (Table S2). These results were in line with those produced from Mäule staining of thin sections of lines #13 and #21, showing stronger reduction of red coloration compared to the wild type and other *HpSK* lines, suggesting a decrease in the amount of S units in lignin of these two lines (Figure S7). Moreover, the total yield of the three monomers with H, G, and S units released by thioacidolysis, which represents the level of uncondensed β–O–4 units in lignin, was reduced in both lines (0.53 and 0.75-fold decrease

compared to the wild type, respectively). The total yield and monomeric composition of thioacidolysis monomers in the other two transgenic lines (#15 and #22) were comparable to those in the wild type.

HSQC spectra of cell wall residues (CWRs) from wild-type and transgenic lines were obtained to assess the frequency of interunit linkages in lignin (Figure 4). This technique helps analyze non-condensed units including β–O–4 substructures (A substructure in Figure 4), in addition to condensed units of lignin quantitatively, such as phenylcoumaran, resinol, and spirodienone structures (B, C, and SD substructures in Figure 4, respectively). The relative abundance of these substructures apparently did not differ between the wild type (Figure 4) and the transgenic lines, even in plants with higher SK activity. A typical characteristic of lignin in lines #13 and #21 revealed by HSQC spectra was an increase in *p*-hydroxybenzoate (Figure 4) residues bound with ester bonds. Note that we were not able to confirm the elevation in H units from these spectra because of a bisphenol-A impurity from the tubes used for milling; see the caption of Figure 4. Therefore, thioacidolysis was more appropriate for determining the H-unit levels expressed relative to the sum of the three units (H+G+S).

In addition to the change in the monomeric composition of lignin, different levels in phenolic compounds that were released by an alkaline treatment were detected in the transgenic lines (Table S3). Compared to the wild type, levels of *p*-hydroxybenzaldehyde increased significantly in lines #13 and #21 with elevated H unit content (29- and 24-fold increase compared to the wild type, respectively). Furthermore, a slight but significant increase in *p*-hydroxybenzoic acid was released from line #13 (supporting the NMR observation, Figure 4B), but not in line #21 ($P = 0.108$).

Effects of *HpSK* expression on metabolites

In order to understand the effects of decreased cytosolic shikimate availability on the biosynthesis of phenylpropanoids and monolignols, we analyzed the changes in the phenolic metabolism. Comparative phenolic profiling was performed using UHPLC–MS of the methanol-soluble fraction prepared from debarked stems of five individual plants of the wild type and of each transgenic line. In total, 19,503 compound ions were detected in the extracted fraction. After removing ions with signals <0.005% of those of the most abundant ions, 7,474 compound ions were analyzed using a principal component analysis (PCA) to visualize general differences between the samples (Figure S8). The PCA plots showed that the two transgenic lines with higher H lignin unit levels (#13 and #21) were separated from the wild type and the other two transgenic lines (#15 and #22; Figure S8a). The intensities of 2,119 compound ions differed significantly between the wild type and lines with higher H lignin unit content. Among the differentially abundant metabolite ions, the levels of 570 and 172 ions increased and decreased, respectively, in both lines #13 and #21, compared to the wild type. In addition, 236 and 24 increased ions and 1,047 and 70 decreased ions were uniquely detected in #13 and #21, respectively.

Metabolites that showed a significant difference in at least one of the transgenic lines #13 and #21, and which were structurally characterized, are listed in Table 2. Shikimate was below the

detection level in all samples. Although the detected signal was too weak for a confident identification by MS/MS, a compound with a high probability of being S3P could be detected in the extracts of the two lines (lines #13 and #21) with high *HpSK* expression, but not in those of the wild type (Table 2). Furthermore, 5-enolpyruvylshikimate-3-phosphate (EPSP), which is synthesized from S3P and phosphoenolpyruvate by endogenous EPSP synthase, also significantly accumulated in lines #13 and #21. In contrast, a significant reduction in caffeoyl shikimate occurred in both transgenic lines, suggesting that transesterification of *p*-coumaroyl-CoA ester with shikimate and/or subsequent 3'-hydroxylation of the resulting shikimate ester was inhibited by *HpSK* expression.

Levels of phenylalanine and tryptophan, which are amino acid metabolite derivatives from the shikimate pathway, were both lower in line #13 than in the wild type. The level of phenylalanine was also lower in line #21. Although no significant differences could be detected at $P = 0.05$, tyrosine showed a trend to be lower in both lines, compared to that in the wild type. In addition to the precursors of secondary metabolites mentioned above, numerous oligolignols with G and S units were decreased in both lines. Among oligolignols identified in the two transgenic lines, three oligolignols with H units, [H(8-O-4)H, H(8-O-4)G', and H(8-O-4)S'] were detected in #13 and #21, while they remained under the detection limit in wild type. In addition, H(8-O-4)S(8-5)G' was significantly increased in line #13, compared to the wild type and showed a tendency to be increased in line #21.

Transcriptional responses of lignin biosynthetic genes to *HpSK* expression

To examine the global changes in gene expression induced by the *HpSK* expression, RNA sequencing was performed using total RNA isolated from developing xylem tissues of the wild type and of three transgenic lines (#13, #15, and #21). Highest *HpSK* expression was detected in line #13 (9,433 ± 879 counts per million, CPM), followed by #21 (7,848 ± 1,061 CPM) and #15 (20 ± 6 CPM; Figure 5). No *HpSK* expression was detected in the wild type. PCA of all annotated genes showed that the high *HpSK*-expressing transgenic lines (#13 and #21) could be separated from the wild type and line #15, based on a combination of PC1 and PC2 (Figure S9).

DESeq2 (ver. 1.24.0; Love *et al.*, 2014) analysis identified 2,464 differentially expressed genes (DEGs) with significant changes (adjusted *p*-value [P_{adj}] <0.05) in #13, 1,563 DEGs in #21, and 207 DEGs in #15, compared to the wild type (Figure S9); 80% of the DEGs (1,394 of 1,749 DEGs) among upregulated genes and 87% of the DEGs (1,615 of 1,859 DEGs) among downregulated genes were identified only in one of the transgenic lines. In lines #13 and #21 with higher *HpSK* expression, 334 and 212 DEGs were shared upregulated and downregulated genes, respectively. Gene Ontology (GO) category enrichment analysis was conducted on the 334 and 212 DEGs to evaluate the transcriptional impact on lines #13 and #21, which revealed that downregulated genes were involved in carbohydrate metabolic processes. The upregulated genes were associated with nitrogen metabolic processes. We examined the expression patterns of genes for monolignol biosynthesis and polymerization. *HpSK* expression did not significantly affect expression levels of the majority of genes associated with lignin biosynthesis (Figure 5

and Figure S10).

Improved saccharification efficiency of wood of *HpSK* transgenic hybrid aspens

Table S4 shows the monosaccharide composition of the cell walls of wild-type and transgenic plants. The levels of each monosaccharide did not differ significantly, except for a slight reduction in mannose levels in the two transgenic lines. The total levels of detected monosaccharides in CWR were comparable between wild-type and transgenic plants. To evaluate the effect of changes in lignin monomeric composition in the transgenic plants on biomass recalcitrance, enzyme saccharification efficiency was measured with or without alkaline pretreatment. Regardless the treatment prior to the saccharification, higher amounts of glucose were released from CWR of both transgenic lines, compared to the wild type, after 24 h enzymatic hydrolysis (Figure 6a). The saccharification efficiency of CWR from the transgenic lines was increased by up to 61% with no pretreatment and by 36% after a 62.5 mM NaOH pretreatment.

Figure 6b shows the lignin content in CWR after the alkaline treatment with 62.5 mM NaOH. The levels of residual lignin in the pretreated CWR were lower in the transgenic lines, compared to the wild type, indicating that lignin in the transgenic plants was easier removed from the cell wall through the pretreatment, even though the initial level of lignin content in wild-type and transgenic plants was comparable.

Discussion

In the present study, *HpSK* was successfully expressed in hybrid aspen. Higher SK activity was observed in crude extracts from stems of lines #13 and #21 (Figures 3 and S2). The amount of shikimate in the cytosol was therefore expected to remain low, which in turn is expected to reduce the metabolic flux via HCT. Poplar HCT can react with hydroxyl residues at the C4 and C5 positions of shikimate, in addition to the C3 position (Kim *et al.*, 2011). Because S3P has two free hydroxy residues at C4 and C5, we tested whether *p*-coumaroyl S3P can be synthesized from externally supplied *p*-coumaroyl-CoA and S3P by endogenous HCT in crude extracts from a wild-type plant. However, no putative *p*-coumaroyl S3P peak was detected in the reaction mixture, and *p*-coumaroyl shikimate remained barely detectable until 20 min reaction time (Figure S4). Even though a small amount of *p*-coumaroyl shikimate was detected at 120 and 600 min, it may have been produced by the use of shikimate that was derived from the externally supplied S3P upon dephosphorylation by an endogenous phosphatase. In contrast, in crude protein extracts of lines #13 and #21, S3P was continuously produced from shikimate in the presence of ATP, a coenzyme of SK (Figure S2). These results suggest that endogenous HCTs in crude extracts cannot use S3P as a substrate, and endogenous phosphatase activity against S3P is markedly lower than SK activity against shikimate. Considering these results as well as data from the metabolomic analysis, we can conclude that *HpSK* expression efficiently reduces cytosolic shikimate availability.

The higher SK activity led to a growth reduction in the transgenic lines, #13 and #21 (Figure 2 and Table S1). Lignin modified plants including *HCT*-, *C3'H*-, and

CSE-downregulated plants were also dwarfed (Franke *et al.*, 2002b, Besseau *et al.*, 2007, Sykes *et al.*, 2015, de Vries *et al.*, 2021a). The mechanism of the growth reduction associated with lignin modifications has not been fully elucidated (Ha *et al.*, 2021). One of the possible reasons is the collapse of xylem vessels that impairs water and solute transport in the dwarfed plants. The growth phenotype could sometimes be recovered by vessel-specific complementation of the downregulated genes (Petersen *et al.*, 2012, Kim *et al.*, 2014, Vargas *et al.*, 2016, De Meester *et al.*, 2018). However, no visible collapse was observed in xylem vessels of lines #13 and #21 (Figure S7). Although accumulation of flavonoids (Besseau *et al.*, 2007), ferulic acid (Xue *et al.*, 2015), and salicylic acid (Gallego-Giraldo *et al.*, 2011) cause dwarfism in the lignin-modified plants, none of these metabolites could be identified in the present metabolomic analysis (Table 2). Auxin-regulated growth is fine-tuned by early steps in phenylpropanoid biosynthesis, and metabolites accumulating upstream of the C4H step such as *t*-cinnamic acid and its photo-isomer *c*-cinnamic acid have also been proposed to contribute to lignin modification induced dwarfism (Steenackers *et al.*, 2017, Kurepa *et al.*, 2018). In the present metabolomic analysis, most of the structurally identified metabolites, including L-phenylalanine that is the direct precursor of *t*-cinnamic acid, decreased in lines #13 and #21 compared to wild type. The growth reduction in the lines might occur as the result of an imbalance in phenylpropanoid pathway homeostasis.

The reduction of cytosolic shikimate availability by *HpSK* expression in the hybrid aspen was expected to produce a similar effect on lignin composition as downregulation or knockout of *HCT*, *C3'H*, or *CSE*. In fact, a significant increase in H unit levels and a decrease in both G and S units in lignin occurred in lines #13 and #21 (Table 1, Figure S6, and Table S2), as was previously observed in transgenic poplars downregulated for *HCT*, *C3'H*, and *CSE* and in poplar mutants deficient in *HCT1* and in both *CSE1* and *CSE2* (Coleman *et al.*, 2008, Vanholme *et al.*, 2013a, Peng *et al.*, 2014, Saleme *et al.*, 2017, Kim *et al.*, 2020, de Vries *et al.*, 2021a). Comparable to *C3'H*-downregulated poplars (Ralph *et al.*, 2006, Kim *et al.*, 2020), lines #13 and #21 showed lower monomer yields (i.e., the sum of H, G, and S monomers) after thioacidolysis. Significantly reduced thioacidolysis monomer yields have also been reported in *HCT*-downregulated tobacco and loblolly pine (Hoffmann *et al.*, 2004, Wagner *et al.*, 2007), but not in *CSE1*- and *CSE2*-downregulated poplars (Saleme *et al.*, 2017). The lower monomer yields in lines #13 and #21 indicate that a lower fraction of units is coupled solely via β–O–4 linkages in these transgenic lines as compared to wild type. Our NMR analysis showed no differences in the overall fraction of β–O–4 linkages between the different lines (Figure 4). The combined observations (thioacidolysis monomer yield and the fraction of β–O–4 as determined via NMR) hint that the distribution of the β–O–4 linkages over the lignin polymer has been changed in the transgenic lines, in such a way that a similar fraction of β–O–4 linkages gives rise to fewer released monomers as compared to the wild type – NMR in principle measures the total β-ether units whereas thioacidolysis monomers arise only from units that are linked by two β-ethers (β–O–4 and 4–O-β, or just one, β–O–4, in the case of a phenolic end-unit). The higher level of H units in the lignin of the two transgenic lines might contribute to the lower monomer

yield, as also shown in the *C3'H*-downregulated poplar (Ralph *et al.*, 2006).

Furthermore, *p*-hydroxybenzoate bound to lignin increased in lines #13 and #21, as is the case in *HCT*- and *C3'H*-downregulated poplars (Ralph *et al.*, 2012, Peng *et al.*, 2014, Kim *et al.*, 2020), but not in *CSE*-downregulated and mutated poplars (Saleme *et al.*, 2017, de Vries *et al.*, 2021a). Goacher *et al.* (2021) reported that *p*-hydroxybenzoate specifically decorates the lignin of fibers in poplar xylem. *p*-Hydroxybenzoylated lignins arise from γ-*p*-hydroxybenzoylated monolignols synthesized before being transferred to the cell wall where they are used for lignin polymerization (Morreel *et al.*, 2004, Lu *et al.*, 2015). These results are in line with those of previous reports suggesting that the level of *p*-hydroxybenzoate was negatively correlated with the amount of S units in lignin of natural poplar variants (Yoo *et al.*, 2018) and of a *F5H*-overexpressing transgenic poplar (Stewart *et al.*, 2009); however, the *p*-hydroxybenzoate ester occurs almost exclusively on S-lignin units in poplar (Regner *et al.*, 2018). Recently, an acyltransferase gene (*PHBMT1*, Potri.001G448000) responsible for biosynthesis of γ-*p*-hydroxybenzoylated monolignols has been identified in poplar (de Vries *et al.*, 2021b, Zhao *et al.*, 2021). Overexpression of *PHBMT1* in the transgenic poplar led to an increase in *p*-hydroxybenzoylated lignin units and changes in solvent dissolution rate of the lignin in cell walls. Although *p*-hydroxybenzoate bound to lignin increased in lines #13 and #21, there was no significant difference in the expression of *PHBMT1* between wild type and the two *HpSK* lines. Most likely, the levels of *p*-hydroxybenzoate increased in lines #13 and #21 because of the higher availability *p*-coumaroyl-CoA, a proposed C6-C3 precursor of the C6-C1 *p*-hydroxybenzoyl-CoA (Loscher and Heide, 1994).

A further characteristic of the *HpSK* lines was a significant increase in *p*-hydroxybenzaldehyde released by alkaline treatments of CWR (Table S3). Cell wall-bound *p*-hydroxybenzaldehyde was also elevated in an *Arabidopsis ref8* mutant deficient in *C3'H* (Franke *et al.*, 2002a). The aldehyde may be mainly derived from β-*p*-hydroxybenzaldehyde ether and/or β-*p*-coumaraldehyde ethers in lignin, made from incorporation of *p*-hydroxybenzaldehyde and *p*-coumaraldehyde monomers in the lignin, respectively. Both monomers are biosynthesized independent from the HCT reaction. *p*-Coumaraldehyde is made from a single conversion from *p*-coumaroyl-CoA via CCR (Figure 1). The exact route towards *p*-hydroxybenzaldehyde is not known, but is likely to go via *p*-coumaraldehyde and *p*-hydroxybenzoyl-CoA. The observed increase in *p*-hydroxybenzaldehyde released by alkaline treatments of *HpSK* CWR, is thus in line with the reduced availability of shikimate and therefore the reduced flux through the HCT-catalyzed reaction in these *HpSK* lines.

Unlike transgenic poplars with suppressed or deficient *HCT*, *C3'H*, and *CSE*, no reduction in lignin content was observed in lines #13 and #21 (Table 1). Our data, obtained with the acetyl bromide method, were confirmed by the cysteine–assisted sulfuric acid (CASA) method (Lu *et al.*, 2021). The CASA procedure can be applied to materials with altered lignin composition and provides results that are well correlated to the total lignin contents measured using the Klason method, the most popular procedure for lignin quantification. In fact, apart from a few exceptions, repression of any of the monolignol biosynthetic genes (from *C4H* to *CAD*) typically results in a

significant decrease in lignin content in transgenic woody plants (Chanoca *et al.*, 2019). Although shikimate is considered to be a good acyl acceptor in the reaction catalyzed by HCT (Levsh *et al.*, 2016), a poplar HCT (HCT6), as well as other plant HCTs, exhibit broad substrate promiscuity *in vitro* against various acceptors such as quinate and catechol (Eudes *et al.*, 2016). Therefore, we cannot rule out the possibility that such metabolites were used in place of shikimate as acceptors for the HCT to support partially the biosynthesis of caffeoyl-CoA and its downstream metabolites to maintain the normal lignin level in *HpSK* lines. As indicated in Figure S2, HCT activity was still detected in the extracts of lines #13 and #21. Our results suggest that the conversion of *p*-coumaroyl-CoA to *p*-coumaroyl-shikimate and subsequently to downstream metabolites such as caffeoyl shikimate and caffeoyl-CoA progresses more slowly in *HpSK*-expressing hybrid aspen. The significant decrease and/or downward trend in the level of monolignols, oligolignols containing G and S units, and their precursor (caffeoyl shikimate), observed in lines #13 and/or #21, may reflect the delayed lignin biosynthesis. This delay is probably insufficient to decrease the lignin content in both lines.

Most of the identified metabolites with a significantly different abundance were decreased in the higher *HpSK*-expressing lines, whereas hexose derivatives of dihydroferulic and sinapic acids were increased in these lines (Table 2). The latter compounds are potentially synthesized via direct hydroxylation of *p*-coumarate by *p*-coumarate 3-hydroxylase (C3H, also known as a cytosolic ascorbate peroxidase; Figure 5). Barros *et al.* (2019) reported that, in addition to C3'H, C3H is also involved in lignin biosynthesis and provides redundancy to the early steps of the monolignol biosynthetic pathway both in dicotyledonous (*Arabidopsis*) and monocotyledonous (*Brachypodium*) plants. Expression of a putative gene for C3H (*APX-C3H2*, Potri.016G084800) was slightly but significantly upregulated in the higher *HpSK*-expressing lines (Figure 5). The increased expression of *APX-C3H2* potentially alleviated the supply limitation of G- and S-type precursors via the HCT- and subsequent C3'H-catalyzed steps, and might also contribute to maintaining the lignin content of *HpSK* lines at the same level as in the wild type.

In addition to the decrease in metabolites of the general phenylpropanoid and monolignol biosynthetic pathways, and in oligolignols, the levels of phenylalanine and/or tryptophan, which are downstream metabolites of the shikimate pathway, were also reduced in lines #13 and #21 (Table 2). In addition, tyrosine accumulation is likely lower in the transgenic lines than in the wild type. In contrast, EPSP accumulation was significantly increased in both lines. It was proposed that feedback regulation of enzymes acting in the shikimate pathway is mediated by metabolites such as aromatic amino acids (Maeda and Dudareva, 2012). For example, enzymes such as arogenate dehydratase and chorismate mutase are inhibited by phenylalanine. Yokoyama *et al.* (2021) also reported that, in *Arabidopsis*, caffeate and caffeoyl shikimate, in addition to *p*-coumaroyl shikimate, can inhibit the activity of three isoforms of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, the first enzyme of the shikimate pathway. These results suggest that the metabolic changes caused by *HpSK* expression affect cytosolic shikimate availability as well as the shikimate pathway itself, even though the gene expression levels of arogenate dehydratase, chorismate mutase, and endogenous SK did not

differ significantly from those in the wild type. Feedback regulation of the shikimate pathway enzymes by reducing downstream metabolites such as phenylalanine and caffeoyl shikimate has the potential to contribute, at least partially, to preventing excessive inhibition of the pathway flux and thus maintaining the normal level of lignin in the *HpSK* lines.

CWRs of *HpSK* lines exhibited higher saccharification efficiency both with and without alkaline pretreatment. A slight modification in monosaccharide composition, and no significant differences in total monosaccharides and total lignin in lines #13 and #21, suggest that the improved efficiency was due to structural modifications of lignin in these plants. Unlike *HpSK* lines, most transgenic trees with improved saccharification efficiency exhibit not only changes in the monomeric composition but also show markedly or slightly reduced lignin content (Chanoca *et al.*, 2019). It has been reported that reduction of lignocellulose recalcitrance is also partially related to the molecular weight of lignin in the cell wall. Suppression of *C3'H* or *HCT* expression in transgenic alfalfa was reported to cause a decrease in the molecular weight of lignins as well as increases in the H unit of lignin and in saccharification yields (Ziebell *et al.*, 2010, Tolbert *et al.*, 2014, Meng *et al.*, 2017). In contrast with the latter transgenic plants, there were no obvious differences in the molecular weight distributions of isolated lignin prepared from *HpSK* lines compared to that prepared from the wild type (Figure S5), indicating that the improved saccharification efficiency and increased extractability of lignin (Figure 6) are not caused by the change in lignin molecular weight in these lines. Despite its higher lignin content, the H-lignin dominant mutant of *A. thaliana* exhibited higher saccharification yield compared to the wild type (Shi *et al.*, 2016), suggesting that the improved saccharification efficiency observed in the present study could be closely associated with an increase in the H units of lignins in *HpSK* lines.

Experimental procedures

Plant material

A *HpSK* gene of *H. pylori* was introduced into hybrid aspen (*P. tremula* × *P. tremuloide*s, cv. T89) under the control of the hybrid aspen *C4H* promoter (*PkC4Hpro*) (Kawai *et al.*, 1996) and a heat shock protein terminator (Sakamoto *et al.*, 2020). Transgenic plants were produced as described previously (Sakamoto *et al.*, 2016). Wild-type and transgenic plants were propagated vegetatively in agar medium under sterilized conditions and were then grown in pots at 25 °C with a 16/8 h light/dark cycle under fluorescent light in a climate-conditioned room.

Semiquantitative RT-PCR and RNA-seq analyses

Debarked stems were collected from 15-week-old plants and were immediately frozen in liquid nitrogen. Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). First-strand cDNA was synthesized using a Transcriptor First Strand cDNA Synthesis Kit (Roche Diagnostics K.K., Tokyo, Japan) for semiquantitative RT-PCR, and DNA derived from *HpSK* transcripts was amplified using GoTaq DNA Polymerase (Promega K.K., Tokyo, Japan). Transcripts of the 18S ribosomal RNA gene were amplified as a positive control. PCR was

performed under the following thermocycling conditions: 95 °C for 3 min followed by 30 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min, with a final step at 72 °C for 5 min. Sequences of the RT-PCR primers are shown in Table S5.

For RNA-seq analysis, after quality inspection of total RNA using an Agilent 2100 Bioanalyzer (Agilent Technologies Japan, Ltd., Tokyo, Japan), an RNA-seq library was constructed using a TruSeq Stranded mRNA library kit (Illumina, San Diego, CA, USA) with three independent biological replicates of the wild type and of three transgenic lines (#13, #15, and #21). The libraries were sequenced on a NovaSeq 6000 platform (Illumina) which resulted in 42.9–65.8 million paired-end reads of 101 bp per sample. RNA-seq was performed by Macrogen Japan (Tokyo, Japan). The sequence data were deposited in the DDBJ Sequence Read Archive under accession number DRA012486.

Raw reads were trimmed using Trimmomatic software (ver. 0.3.9) (Bolger *et al.*, 2014) with the following parameters: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:36. For transcriptome profiling of the native gene of aspen, the trimmed paired reads were mapped to the reference genome of *Populus tremula* v2.2 (ftp://plantgenie.org/Data/PopGenIE/Populus_tremula/v2.2/fasta/Potra02_genome.fasta.gz) using HISAT2 (ver. 2.2.1) (Kim *et al.*, 2019) with the following options: --no-discordant and --no-mixed. The obtained Sequence Alignment Map was converted into a Binary Alignment Map using samtools (ver. 0.3.3) (Li *et al.*, 2009). Reads were counted using HTSeq (ver. 0.12.4) (Anders *et al.*, 2015). Raw counts of all samples were merged into a file and were processed using R software (ver. 3.6.0, R Core Team). For transcriptome profiling, we performed a PCA with counts per million reads for all samples of wild-type and transgenic lines. Comparing wild-type and transgenic lines, we detected DEGs using the DESeq2 package (ver. 1.24.0) (Love *et al.*, 2014) of R. The number of DEGs, average counts per million reads, fold change (FC), and Benjamin and Hochberg-adjusted P_{adj} were documented. Similarities and differences of upregulated and downregulated DEGs are shown in Venn diagrams. GO enrichment analysis of the DEGs was performed using the g:Profiler Web server (https://biit.cs.ut.ee/gprofiler/) (Raudvere *et al.*, 2019).

For evaluation of *HpSK* transgene expression, trimmed reads were pseudo-aligned using kallisto (Bray *et al.*, 2016) and the *HpSK* sequence. The counts were adjusted with the number of trimmed reads to produce counts per million reads. The sequencing data were deposited in the DDBJ under the accession no. DRA012486.

Enzymatic activity of SK in stems

Debarked xylem tissue (100 mg) prepared from plants grown for 15 weeks after transplantation to pots was ground in liquid nitrogen, and crude enzyme was extracted from ground tissue using 0.5 mL 100 mM Tris–HCl (pH 7.5) for 30 min at 4 °C under rotation (10 rpm). After pelleting cell debris by centrifugation (19,000 × g) for 20 min at 4 °C, the supernatant was subjected to ultrafiltration using an Amicon Ultra centrifugal filter unit (10 kDa cutoff; Merck Millipore, Burlington, MA, USA) to remove endogenous low-molecular-weight compounds including

cofactors. Protein concentrations were determined using the Bradford method (Bradford, 1976). For evaluating SK activity, ultrafiltered crude enzyme (2 mg/mL protein) was incubated with 0.5 mM shikimate in the assay solution (400 µL total volume) containing 100 mM Tris–HCl (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, and 1 mM ATP at 30 °C. A reaction assay without crude enzyme and purified HpSK (2 mg/mL) was used as a control. Portions of the reaction mixtures were periodically collected, and the reaction was stopped by adding the 9-fold volume of 0.1% formic acid. SK activity was expressed in moles of S3P produced per min per milligram of protein. The amounts of shikimate and S3P were measured using UHPLC–MS with an ACQUITY UPLC system (Waters, Milford, MA, USA) coupled with an ACQUITY TQ detector (Waters) and a BEH C18 (2.1 mm ID × 100 mm length, 1.7 µm particle size; Waters) as previously described (Fukuhara et al., 2010). The mobile phase of the UHPLC system was water containing 1% acetonitrile and 0.1% formic acid at a flow rate of 0.35 mL min^{-1}, and temperature was maintained at 30 °C. shikimate and S3P were detected by in the selected ion recording mode with negative ions of *m*/*z* 173 and 253, respectively. The following retention times were used: shikimate, 0.95 min; S3P, 0.93 min. Averages ± standard deviations of three biological replicates were calculated.

Microscopic observation

Stem parts with the same diameter were sampled from plants grown for 15 weeks after transplantation to pots and were fixed in FAA solution (50% ethanol, 10% formaldehyde, and 5% acetic acid). Thin stem sections (20 µm thick) were prepared using a microtome (model MTH-1, Nippon Medical & Chemical Instruments Co., Osaka, Japan). For Mäule staining, sections were stained using 1% $KMnO_4$ for 5 min, after which they were washed three times using distilled water. After two incubation steps in 10% HCl, the sections were washed three times using distilled water and were transferred to a solution of 1.5 M Na_2CO_3. The sections were examined using a microscope (model BZ-X810, Keyence Corporation, Osaka, Japan).

Sub-cellular localization of HpSK in mesophyll protoplasts

To observe the subcellular localization of HpSK in protoplast cells, the expression plasmid for GFP-fused HpSK protein was prepared by the insertion of HpSK PCR fragments amplified using PCR primers (HpSK_GFP_F1 and HpSK_GFP_R1, Table S5) into p35SGFPNHSPG vector (Fujiwara *et al*. 2014) or HpSK_GFP_F2 and HpSK_GFP_R2 into p35S_GFP_HSPG vector. The p35S_GFP_HSPG vector was constructed by inserting the HSP terminator of the SacI and NotI fragment prepared from p35SHSPG (Ho *et al.*, 2021) into p35SGFPG (Oshima *et al.*, 2011). The resultant plasmid was introduced into the *Arabidopsis* mesophyll protoplasts prepared as described previously (Yoshida *et al.*, 2013, Sakamoto *et al.*, 2016), and GFP-fluorescence was then observed using fluorescence microscopy (BZ9000 Generation II Biorevo microscope, Keyence Corporation, Osaka, Japan).

Lignin content and composition

Plants grown for 15 weeks after transplantation to pots (five biological replicates) were debarked, dried in the dark at room temperature, and were cut into small pieces, followed by grinding to particles smaller than 150 μm using a milling machine (Multi-Beads Shocker MB1200C, Yasui Kikai Corporation, Osaka, Japan). To prepare CWR, the ground powder was extracted sequentially using sonication for 30 min with 50 mM NaCl, water (three times), methanol, acetone, ethanol (twice), ethanol/toluene (1:2, v/v, twice), and acetone. Lignin content in CWR (5 mg) was determined by the acetyl bromide (Hatfield *et al.*, 1999) and CASA (Lu *et al.*, 2021) methods.

Monomeric composition of lignin in CWR (7 mg) was analyzed by thioacidolysis. Conventional thioacidolysis products were quantified by gas chromatography flame-ionization detection (GC/FID) after trimethylsilylation (Yue *et al.*, 2012, Sakamoto *et al.*, 2020). For determination of lignin content and composition, five biological replicates were analyzed per plant line.

Py-CG–MS analysis of CWR (0.2 mg each) was performed as described by Sakamoto *et al.* (2020). Data were analyzed as descrived by van Erven et al. (2017). To eliminate the influence of bisphenol-A contamination in CWRs during the grinding process (see caption in Figure 4), new plantlets of wild type, and of lines #13 and #21 (four plants of each line) were grown in pots for 10 weeks in a chamber maintained at 25 °C, and fresh CWRs were prepared for Py-CG–MS analysis. Grinding of wood samples for CWR preparation was performed using stainless steel tubes instead of plastic tubes.

Analysis of cell-wall-bound phenolics

Cell wall-bound phenolics were quantified after their release through alkaline hydrolysis of the CWR (30 mg) and were analyzed subsequently by GC/FID as described previously (Sakamoto *et al.*, 2020).

^{1}H–^{13}C heteronuclear single-quantum coherence (^{1}H–^{13}C HSQC) analysis

CWR prepared from plants grown in soil for 15 weeks was used for nuclear magnetic resonance (NMR) analysis. Samples for 2D NMR were prepared as described previously (Kim and Ralph 2010). Briefly, dried stem wood was (unfortunately – see caption to Figure 4) pre-ground in polycarbonate tubes using a milling machine (Multi-Beads Shocker MB1200C) and was extracted using distilled water and organic solvents successively under ultra-sonication. Extractive-free samples were ball-milled and were then subjected to enzymatic digestion (at 35 °C for 72 h, two times) using Cellulysin cellulase of *Trichoderma viride* Pers. (Merck KGaA, Darmstadt, Germany) in acetate buffer (pH 5). Enzyme lignins containing the residue from enzymatic digestion were dissolved in DMSO-d_6/pyridine-d_5 (4:1) after washing and drying. NMR spectra were produced using a Bruker Biospin Neo 700-MHz spectrometer (Bruker BioSpin Co., Billerica, MA, USA) fitted with a cryogenically cooled 5-mm quadruple-resonance ^{1}H/^{31}P/^{13}C/^{15}N QCI gradient probe with inverse geometry. ^{1}H–^{13}C HSQC experiments were carried out as reported previously (Kim and Ralph, 2010). Relative quantification of aromatic S

and G units (Figure 4) is via uncorrected volume-integration of the contour peaks for S2/6 and G2, and logically dividing the S integral by 2 because it represents two C/H-pairs in one S unit, i.e., Σ(**S**+**G**) = ½**S**2/6 + **G**2 = 100%. Such quantification is considered to be reasonably accurate, but *p*-hydroxybenozate (***p*B**) endgroups are over quantified because of their NMR relaxation properties – they are therefore reported on the Σ(**S**+**G**) basis (Kim and Ralph, 2010, Mansfield *et al.*, 2012). Relative quantification of sidechain units, designated via their characteristic inter-unit linkages [**A** = β–O–4, **B** = β–5 (phenylcoumaran), **C** = β–β (resinol), and **SD** = β–1 (spirodienone)] is via volume-integration of the α-C/H correlation peaks of each unit, is expressed on the basis of the sum of these units measured [Σ(**A**+**B**+**C**+**SD**) = **A**α + **B**α + ½**C**α + **SD**α 100%], and again logically dividing the integral from **C**α by 2 (Kim and Ralph, 2020; Mansfield et al., 2012).

Preparation of dioxane lignin and analysis by gel-permeation chromatography (GPC)

To remove the cellulosic material, cell wall residue, weighing 500 mg, was hydrolyzed by cellulase and cellobiase (see Saccharification assay in Experimental procedures) and washed three times with water. The residue was then suspended in 100 mL of a mixture of dioxane and water (9:1, v/v) containing 0.2 M HCl. The suspension was boiled for 4 h under reflux water. The cooled reaction mixture was filtered using a Büchner funnel and the residue was washed three times with 10 mL of a fresh mixture of dioxane and water. All the filtrates containing the dissolved lignins were pooled and the pH of the resulting solution was adjusted to 3 to 4 by $NaHCO_3$ solution. The solution was concentrated to ~10 mL by rotary evaporation at 45 °C. The concentrated solution was then injected into 50 mL of cold water. The lignin precipitate was recovered by centrifugation (2,000 g at 10 °C for 10 min), washed with pure water, centrifuged again, and dried to recover a dioxane lignin. The isolated dioxane lignin was acetylated with acetic anhydride in pyridine at room temperature and then analyzed by GPC using an HPLC system (Prominence, Shimazu Corporation, Kyoto, Japan) equipped with two GPC columns (2 × Shodex KF-805L, SHOWA DENKO K.K., Tokyo, Japan) and a UV detector (280 nm); Eluent, tetrahydrofuran; Flow rate: 0.5 mL/min. Polystyrene standards and acetylated Kraft lignin were used as references.

Metabolomic analyses

Debarked stems of plants (five biological replicates of each line) grown for 15 weeks in soil were placed in liquid nitrogen and were stored at -80 °C until further utilization. For metabolic analyses, frozen stems were pulverized using a Multi-Beads Shocker (Yasui Kikai Corporation), and approximately 100 mg (20–60 mg dry weight) frozen stem powder was mixed with 1 mL methanol and was shaken (160 rpm) for 15 min at 70 °C. The samples were then centrifuged at maximum speed for 3 min. After transferring the methanol to a new microtube, the solution was freeze-dried. Dried ethanol extracts were re-dissolved in 100 µL cyclohexane/water (1/1), and 80 µL water phase was used for UHPLC–MS analysis.

UHPLC was performed using an ACQUITY UPLC I-Class system (Waters) comprising a binary pump, a vacuum degasser, an autosampler, and a column oven. Chromatographic

separation was carried out on an ACQUITY UPLC BEH C18 (150 × 2.1 mm, 1.7 μm) column (Waters), and temperature was maintained at 40 °C. A gradient of two buffers was used: buffer A (99:1:0.1 water: acetonitrile: formic acid, pH 3) and buffer B (99:1:0.1 acetonitrile: water: formic acid, pH 3) as follows: 99% A for 0.1 min, decrease to 50% A until 30 min, decrease to 0% from 30 to 40 min. The flow rate was 0.35 mL min^{-1}, and the injection volume was 10 μL.

The UHPLC system was coupled to a Vion IMS QTOF hybrid mass spectrometer (Waters). The LockSpray ion source was operated in negative electrospray ionization mode. The collision energy for data dependent acquisition (DDA) MSMS settings were: low mass ramp start 15eV; low mass ramp end: 30eV; high mass ramp start: 30eV; high mass ramp end: 70eV. Nitrogen was employed as desolvation and cone gas. Leucin enkephalin (250 pg μL^{1} solubilized in water:acetonitrile 1:1 [v/v], with 0.1% formic acid) was used for the lock mass calibration, with scanning every 2 min at a scan time of 0.1 s. Profile data were recorded through a UNIFI Scientific Information System (Waters). System (Waters). The raw chromatograms are available from the MetaboLights database repository as study MTBLS3863 (www.ebi.ac.uk/metabolights/MTBLS3863) (Haug *et al.*, 2020).

The data was analysed with Progenesis QI software version 2.4 (Waters). All 19,503 peaks (mass-to-charge ratio [m/z] features) were integrated in the chromatograms of the wild type and four transgenic lines. The 7,474 peaks with an abundance of at least 0.005% of the highest average in the group with the highest peak abundance were selected for further analysis. After analysis of variance, peaks with a *P*-value <0.01 (false discovery rate-adjusted) and a two-fold difference in abundance between transgenic plants and wild type were considered significantly different. PCA and volcano plots were generated using MetaboAnalyst 4.0 (Chong *et al.*, 2018). Compounds were structurally characterized based on MS/MS fragmentation spectra.

Saccharification assay

CWR (10 mg) was mixed with 400 μL NaOH (62.5 mM) and was heated to 120 °C for 1 h. Then, the mixtures were neutralized using 1 M HCl. After centrifugation, the pretreated CWR was washed three times using 1 mL Milli-Q water. For saccharification, the samples were hydrolyzed using 1.8 mL 5 mM sodium citrate buffer (pH 4.8) containing 200 μg/mL sodium azide, 2.8% w/w cellulase (C2730, Sigma Aldrich Japan, Tokyo), and 1.6% w/w cellobiase (C6105, Sigma Aldrich Japan). After incubation at 50 °C under shaking (160 rpm) for 24 h, the samples were centrifuged at 20,000 × g for 10 min, and 20 μL of the supernatant was used for measurement of released glucose using a Glucose CII test kit (Fujifilm Wako Pure Chemical corporation, Osaka, Japan). Absorbance of the samples was measured using a spectrophotometer at 504 nm. Glucose was used to produce a calibration curve.

Acknowledgements

This study was partially supported by the Japan Science and Technology Agency (Advanced Low Carbon Technology Research and Development Program, grant no. JPMJAL1107) and by the Japan Society for the Promotion of Science. SL, HK, and JR gratefully acknowledge funding from the US Department of Energy (DOE) through the Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409). WB acknowledges funding from iBOF project Next-Bioref. We thank Prof. Fengxia Yue (South China University of Technology) for providing a chemical compound used for thioacidolysis product authentication and Mr. Zen Ookawa (AIST) for providing technical support in vector construction and transient gene expression in protoplast cells.

Author contributions

NK, SS, EM, NM, and SK conceived the study. NK and SS prepared the gene construct for genetic transformation, and MU generated the transgenic hybrid aspens. SH and NK measured enzymatic activity. SH (Shi Hu), SL, YM, HK, and JR performed the lignin analysis. SN and NT analyzed gene expression by RNA sequencing. GG performed metabolomic analyses, and GG and RV interpreted the metabolite data. SH and SS analyzed the sugar composition and saccharification efficiency. GPC analysis of the isolated lignin and subcellular localization analysis of GFP-fused HpSK protein were performed by Shjiro Hishiyama and SS, respectively. SH performed the other experiments. SH and SK wrote the initial draft, and all other authors contributed to the revisions.

Conflict of interest

The authors declare no conflicts of interest.

Data availability statement

All relevant data supporting the results presented in this work are available within the article and the supporting information.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Figure S1. Sub-cellular localization of green fluorescent protein (GFP)-fused HpSK in an *Arabidopsis* mesophyll protoplast. Chimeric constructs encoding GFP fused to the amino- (GFP-HpSK) and carboxyl-termini (HpSK-GFP) of HpSK were expressed transiently in the cells under the control of the Cauliflower Mosaic Virus 35S promoter. GFP fluorescence derived from the fusion proteins was localized predominantly in the cytoplasm as in the case of GFP expressed alone. DIC, differential interference contrast microscopic observation. Bars indicate 30 μm.

Figure S2. Time-dependent consumption of shikimate and formation of S3P in crude extracts of wild-type plants (a) and transgenic lines #13 and #21 (b and c, respectively). Ultrafiltered crude extracts (10 kDa cut, 2 mg/mL protein equivalent) were incubated with 0.5 mM shikimate and 0.5 mM *p*-coumaroyl-CoA in the presence of 1 mM ATP, 5 mM $MgCl_2$, and 50 mM KCl in 100 mM Tris-HCl (pH 7.5) at 30 °C. Compound I was also detected in the same reaction mixtures, and it was identified as *p*-coumaroyl shikimate, as indicated in Figure S3. The level of *p*-coumaroyl shikimate synthesized in the same reaction mixture with crude extracts from the transgenic lines (b, c) was lower than that produced from the wild type (a) 120 min after the reaction.

Figure S3. Identification of compound I as *p*-coumaroyl shikimate by HPLC–MS analysis. Shikimate and *p*-coumaroyl-CoA were reacted with recombinant HCT purified from *Escherichia coli* expressing *HCT* originating from *Arabidopsis thaliana*. Reaction products were monitored using a photodiode array and tandem quadrupole mass detectors.

Figure S4. Reaction of S3P and *p*-coumaroyl-CoA in the crude extract of wild-type hybrid aspens. Ultrafiltered crude extract (10 kDa cut, 2 mg/mL protein equivalent) was incubated with 0.5 mM S3P and 0.5 mM *p*-coumaroyl-CoA in 100 mM Tris-HCl (pH 7.5) at 30 °C. Shikimate was detected 20 min after the start of the reaction, indicating that endogenous phosphatase catalyzes dephosphorylation of S3P in the reaction mixture. Synthesis of *p*-coumaroyl shikimate was observed 120 min after the reaction, suggesting that the resulting shikimate reacts with *p*-coumaroyl-CoA under the catalytic action by endogenous HCT.

Figure S5. Gel permeation chromatography of isolated dioxane lignins. Elution curves of lignins from different lines, Kraft lignin (Sigma-Aldrich, Tokyo, Japan), and polystyrene standards with six different molecular weights (A, 1,730,000; B, 523,000; C, 60,450; D, 9,920; E, 1,990; F, 580) are shown. The sharp peaks (asterisk) observed in the chromatograms other than that of Kraft lignin originate from bisphenol-A contamination in the plastic tubes used for the milling of cell wall residues (CWRs).

Figure S6. Partial GC-MS traces showing the detection of conventional thioacidolysis products derived from lignin units linked with β-O-4 bond. Peaks labeled with H, G, and S correspond to the thioacidolysis products with *p*-hydroxyphenyl, guaiacyl, and syringyl unit, respectively. Total-ion (TIC) and selected-ion (m/z 239, 269, and 299) chromatograms are indicated with different colors. The selected-ion chromatograms are shown at 5-fold magnification and the y-axis zero-point is offset so that the chromatograms are more clearly resolved and viewed. IS indicates a peak derived from an internal standard (tetracosane).

Figure S7. Mäule staining of thin stem sections prepared from wild-type and four transgenic lines.

Scale bars indicate 100 μm.

Figure S8. Metabolome analysis of methanol-soluble extracts in stems. (a) Plot of principal component analysis (PCA) on 7,474 peaks (excluding variables with missing values and normalized by log-transformation, pareto scaled). PC1 and PC2 account for 48% and 9% of the total variance, respectively. Ellipses indicate 95% confidence regions. Pink dots: wild type; red dots: line #13; green dots: line #15; purple dots: line #21; light blue dots: line #22. (b) Venn diagrams of the numbers of peaks with significantly differential intensities between wild type and line #13, and wild type and line #21. Five independent biological samples were used.

Figure S9. Scatter plots of PCA scores derived from stem transcriptome and Venn diagram of significantly up-regulated and down-regulated genes (P_{adj} <0.05) of transgenic lines #13, #15, and #21 compared to the wild type. (a) PCA score plots of transcriptome data from the stem of the wild type and three transgenic lines. The relationships between PC1 and PC2, and between PC1 and PC3 are presented. (b) Venn diagram of DEGs in comparison between the wild type and the three transgenic lines. RNA isolated from three individual wild-type and transgenic plants was used.

Figure S10. Expression profiles of genes encoding laccase and peroxidase, which are predicted to be involved in lignin polymerization.

Table S1. Growth characteristics of wild-type and transgenic plants. Height, diameter, and dry weight of stems prepared from the wild-type (WT) and transgenic lines. Data represent means ± standard error from five biological replicates. Each sample was measured 15 weeks after transferring the plants to soil. Asterisks indicate significant differences from the wild type according to Student's *t*-test (** $P < 0.01$).

Table S2. Structural characterization of lignin using pyrolysis-gas chromatography–mass spectrometry. Pyrolysis followed by GC–MS analysis was performed with cell wall residues (CWR) as described in previous reports (Yoshida *et al.*, 2013, Sakamoto *et al.*, 2020). The analysis was performed with four biological replicates of each line. Values in bold font indicate significant differences compared to the wild type ($P < 0.05$, Student's *t*-test). Intensity was expressed as the peak area normalized to the CWR used in the analysis.

Table S3. Phenolic compounds released from CWR after alkaline treatment. Data represent means ± standard error from five biological replicates. Asterisks indicate significant differences from the wild type using the Student's *t*-test (* $0.01 < P < 0.05$; ** $P < 0.01$).

Table S4. Monomeric sugar compositions in CWR released from stems of wild-type and transgenic plants after hydrolysis using sulfuric acid. Value represents means ± standard error

from five biological replicates. Asterisks indicate significant differences from the wild type using the Student's *t*-test (*$P < 0.05$).

Table S5. Primers used in RT-PCR and vector construction.

References

Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.

Barros, J., Escamilla-Trevino, L., Song, L., Rao, X., Serrani-Yarce, J.C., Palacios, M.D., Engle, N., Choudhury, F.K., Tschaplinski, T.J., Venables, B.J., Mittler, R. and Dixon, R.A. (2019) 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nat Commun*, **10**, 1994.

Barros, J., Serk, H., Granlund, I. and Pesquet, E. (2015) The cell biology of lignification in higher plants. *Ann Bot*, **115**, 1053-1074.

Bassard, J.E., Richert, L., Geerinck, J., Renault, H., Duval, F., Ullmann, P., Schmitt, M., Meyer, E., Mutterer, J., Boerjan, W., De Jaeger, G., Mely, Y., Goossens, A. and Werck-Reichhart, D. (2012) Protein-protein and protein-membrane associations in the lignin pathway. *Plant Cell*, **24**, 4465-4482.

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B. and Legrand, M. (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell*, **19**, 148-162.

Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. *Annu Rev Plant Biol*, **54**, 519-546.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.

Bonawitz, N.D., Kim, J.I., Tobimatsu, Y., Ciesielski, P.N., Anderson, N.A., Ximenes, E., Maeda, J., Ralph, J., Donohoe, B.S., Ladisch, M. and Chapple, C. (2014) Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. *Nature*, **509**, 376-380.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, **72**, 248-254.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, **34**, 525-527.

Chanoca, A., de Vries, L. and Boerjan, W. (2019) Lignin engineering in forest trees. *Front Plant Sci*, **10**.

Cheng, W.C., Chang, Y.N. and Wang, W.C. (2005) Structural basis for shikimate-binding specificity of Helicobacter pylori shikimate kinase. *J Bacteriol*, **187**, 8156-8163.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S. and Xia, J. (2018)

MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*, **46**, W486-W494.

Coleman, H.D., Park, J.Y., Nair, R., Chapple, C. and Mansfield, S.D. (2008) RNAi-mediated suppression of *p*-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. *Proc Natl Acad Sci U S A*, **105**, 4501-4506.

De Meester, B., de Vries, L., Ozparpucu, M., Gierlinger, N., Corneillie, S., Pallidis, A., Goeminne, G., Morreel, K., De Bruyne, M., De Rycke, R., Vanholme, R. and Boerjan, W. (2018) Vessel-specific reintroduction of CINNAMOYL-COA REDUCTASE1 (CCR1) in dwarfed *ccr1* mutants restores vessel and xylary fiber integrity and increases biomass. *Plant Physiol*, **176**, 611-633.

de Vries, L., Brouckaert, M., Chanoca, A., Kim, H., Regner, M.R., Timokhin, V.I., Sun, Y., De Meester, B., Van Doorsselaere, J., Goeminne, G., Chiang, V.L., Wang, J.P., Ralph, J., Morreel, K., Vanholme, R. and Boerjan, W. (2021a) CRISPR-Cas9 editing of CAFFEOYL SHIKIMATE ESTERASE 1 and 2 shows their importance and partial redundancy in lignification in *Populus tremula* x *P. alba*. *Plant Biotechnol J*, **19**, 2221-2234.

de Vries, L., MacKay, H.A., Smith, R.A., Mottiar, Y., Karlen, S.D., Unda, F., Muirragui, E., Bingman, C., Vander Meulen, K., Beebe, E.T., Fox, B.G., Ralph, J. and Mansfield, S.D. (2021b) pHBMT1, a BAHD-family monolignol acyltransferase, mediates lignin acylation in poplar. *Plant Physiol*, **accepted**.

Eudes, A., Pereira, J.H., Yogiswara, S., Wang, G., Benites, V.T., Baidoo, E.E.K., Lee, T.S., Adams, P.D., Keasling, J.D. and Loque, D. (2016) Exploiting the substrate Promiscuity of hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase to reduce lignin. *Plant Cell Physiol*, **57**, 568-579.

Eudes, A., Sathitsuksanoh, N., Baidoo, E.E.K., George, A., Liang, Y., Yang, F., Singh, S., Keasling, J.D., Simmons, B.A. and Loque, D. (2015) Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnol J*, **13**, 1241-1250.

Franke, R., Hemm, M.R., Denault, J.W., Ruegger, M.O., Humphreys, J.M. and Chapple, C. (2002a) Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of Arabidopsis. *Plant J*, **30**, 47-59.

Franke, R., Humphreys, J.M., Hemm, M.R., Denault, J.W., Ruegger, M.O., Cusumano, J.C. and Chapple, C. (2002b) The Arabidopsis *REF8* gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J*, **30**, 33-45.

Freudenberg, K. and Neish, A.C. (1968) Constitution and Biosynthesis of Lignin. Springer-Verlag, Berlin-Heidelberg-New York.

Gallego-Giraldo, L., Escamilla-Trevino, L., Jackson, L.A. and Dixon, R.A. (2011) Salicylic acid mediates the reduced growth of lignin down-regulated plants. *P Natl Acad Sci USA*, **108**, 20814-20819.

Goacher, R.E., Mottiar, Y. and Mansfield, S.D. (2021) ToF-SIMS imaging reveals that *p*-hydroxybenzoate groups specifically decorate the lignin of fibres in the xylem of poplar and willow. *Holzforschung*, **75**, 452–462.

Ha, C.M., Escamilla-Trevino, L., Yarce, J.C., Kim, H., Ralph, J., Chen, F. and Dixon, R.A. (2016) An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant J*, **86**, 363-375.

Ha, C.M., Rao, X., Saxena, G. and Dixon, R.A. (2021) Growth-defense trade-offs and yield loss in plants with engineered cell walls. *New Phytol*.

Han, C., Zhang, J., Chen, L.L., Chen, K.X., Shen, X. and Jiang, H.L. (2007) Discovery of *Helicobacter pylori* shikimate kinase inhibitors: Bioassay and molecular modeling. *Bioorgan Med Chem*, **15**, 656-662.

Hatfield, R.D., Grabber, J., Ralph, J. and Brei, K. (1999) Using the acetyl bromide assay to determine lignin concentrations in herbaceous plants: Some cautionary notes. *J Agr Food Chem*, **47**, 628-632.

Haug, K., Cochrane, K., Nainala, V.C., Williams, M., Chang, J., Jayaseelan, K.V. and O'Donovan, C. (2020) MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res*, **48**, D440-D444.

Ho, C.K., Bringmann, M., Oshima, Y., Mitsuda, N. and Bergmann, D.C. (2021) Transcriptional profiling reveals signatures of latent developmental potential in *Arabidopsis* stomatal lineage ground cells. *Proc Natl Acad Sci U S A*, **118**.

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

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. and Legrand, M. (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J Biol Chem*, **278**, 95-103.

Holländer-Czytko, H. and Amrhein, N. (1983) Subcellular compartment of shikimic acid and phenylalanine in buckwheat cell suspension cultures grown in the presence of shikimate pathway inhibitors. *Plant Sci Lett*, **29**, 89-96.

Kawai, S., Mori, A., Shiokawa, T., Kajita, S., Katayama, Y. and Morohoshi, N. (1996) Isolation and analysis of cinnamic acid 4-hydroxylase homologous genes from a hybrid aspen, Populus kitakamiensis. *Biosci Biotechnol Biochem*, **60**, 1586-1597.

Kim, B.G., Kim, I.A. and Ahn, J.H. (2011) Characterization of hydroxycinnamoyl-coenzyme A shikimate hydroxycinnamoyltransferase from *Populus euramericana*. *J Appl Biol Chem*, **54**, 817-821.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, **37**, 907-915.

Kim, H., Li, Q.Z., Karlen, S.D., Smith, R.A., Shi, R., Liu, J., Yang, C.M., Tunlaya-Anukit, S.,

Wang, J.P., Chang, H.M., Sederoff, R.R., Ralph, J. and Chiang, V.L. (2020) Monolignol benzoates incorporate into the lignin of transgenic *Populus trichocarpa* depleted in C3H and C4H. *Acs Sustain Chem Eng*, **8**, 3644-3654.

Kim, H. and Ralph, J. (2010) Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d_6/pyridine-d_5. *Org Biomol Chem*, **8**, 576-591.

Kim, J.I., Ciesielski, P.N., Donohoe, B.S., Chapple, C. and Li, X. (2014) Chemically induced conditional rescue of the *Reduced Epidermal Fluorescence8* mutant of Arabidopsis reveals rapid restoration of growth and selective turnover of secondary metabolite pools. *Plant Physiol*, **164**, 584-595.

Kurepa, J., Shull, T.E., Karunadasa, S.S. and Smalle, J.A. (2018) Modulation of auxin and cytokinin responses by early steps of the phenylpropanoid pathway. *Bmc Plant Biol*, **18**, 278.

Levsh, O., Chiang, Y.C., Tung, C.F., Noel, J.P., Wang, Y. and Weng, J.K. (2016) Dynamic conformational states dictate selectivity toward the native substrate in a substrate-permissive acyltransferase. *Biochemistry-Us*, **55**, 6314-6326.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Liang, Y., Eudes, A., Yogiswara, S., Jing, B., Benites, V.T., Yamanaka, R., Cheng-Yue, C., Baidoo, E.E., Mortimer, J.C., Scheller, H.V. and Loque, D. (2019) A screening method to identify efficient sgRNAs in Arabidopsis, used in conjunction with cell-specific lignin reduction. *Biotechnol Biofuels*, **12**, 130.

Liu, Q., Luo, L. and Zheng, L. (2018) Lignins: Biosynthesis and biological functions in plants. *Int J Mol Sci*, **19**.

Loscher, R. and Heide, L. (1994) Biosynthesis of p-hydroxybenzoate from *p*-coumarate and *p*-coumaroyl-coenzyme A in cell-free extracts of *Lithospermum erythrorhizon* cell cultures. *Plant Physiol*, **106**, 271-279.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.

Lu, F., Wang, C., Yue, F. and Ralph, J. (2021) A facile spectroscopic method for measuring lignin content in lignocellulosic biomass. *Green Chem*, **23**, 5106-5112.

Lu, F.C., Karlen, S.D., Regner, M., Kim, H., Ralph, S.A., Sun, R.C., Kuroda, K., Augustin, M.A., Mawson, R., Sabarez, H., Singh, T., Jimenez-Monteon, G., Zakaria, S., Hill, S., Harris, P.J., Boerjan, W., Wilkerson, C.G., Mansfield, S.D. and Ralph, J. (2015) Naturally *p*-hydroxybenzoylated lignins in palms. *Bioenergy Res*, **8**, 934-952.

Maeda, H. and Dudareva, N. (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annual Review of Plant Biology*, **63**, 73-105.

Mansfield, S.D., Kim, H., Lu, F. and Ralph, J. (2012) Whole plant cell wall characterization using solution-state 2D NMR. *Nat Protoc*, **7**, 1579-1589.

Meng, X.Z., Pu, Y.Q., Yoo, C.G., Li, M., Bali, G., Park, D.Y., Gjersing, E., Davis, M.F.,

Muchero, W., Tuskan, G.A., Tschaplinski, T.J. and Ragauskas, A.J. (2017) An in-depth understanding of biomass recalcitrance using natural poplar variants as the feedstock. *Chemsuschem*, **10**, 139-150.

Morreel, K., Ralph, J., Kim, H., Lu, F., Goeminne, G., Ralph, S., Messens, E. and Boerjan, W. (2004) Profiling of oligolignols reveals monolignol coupling conditions in lignifying poplar xylem. *Plant Physiol*, **136**, 3537-3549.

Oshima, Y., Mitsuda, N., Nakata, M., Nakagawa, T., Nagaya, S., Kato, K. and Ohme-Takagi, M. (2011) Novel vector systems to accelerate functional analysis of transcription factors using chimeric repressor gene-silencing technology (CRES-T). *Plant Biotechnol*, **28**, 201-210.

Padmakshan, D., Timokhin, V.I., Lu, F., Schatz, P.F., Vanholme, R., Boerjan, W. and Ralph, J. (2021) Synthesis of hydroxycinnamoyl shikimates and their role in monolignol biosynthesis. *Holzforschung*, **Accepted**.

Peng, X.P., Sun, S.L., Wen, J.L., Yin, W.L. and Sun, R.C. (2014) Structural characterization of lignins from hydroxycinnamoyl transferase (HCT) down-regulated transgenic poplars. *Fuel*, **134**, 485-492.

Petersen, P.D., Lau, J., Ebert, B., Yang, F., Verhertbruggen, Y., Kim, J.S., Varanasi, P., Suttangkakul, A., Auer, M., Loque, D. and Scheller, H.V. (2012) Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnol Biofuels*, **5**, 84.

Ralph, J., Akiyama, T., Coleman, H.D. and Mansfield, S.D. (2012) Effects on lignin structure of coumarate 3-hydroxylase downregulation in poplar. *Bioenergy Res*, **5**, 1009-1019.

Ralph, J., Akiyama, T., Kim, H., Lu, F., Schatz, P.F., Marita, J.M., Ralph, S.A., Reddy, M.S., Chen, F. and Dixon, R.A. (2006) Effects of coumarate 3-hydroxylase down-regulation on lignin structure. *J Biol Chem*, **281**, 8843-8853.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*, **47**, W191-W198.

Regner, M., Bartuce, A., Padmakshan, D., Ralph, J. and Karlen, S.D. (2018) Reductive cleavage method for quantitation of monolignols and low-abundance monolignol conjugates. *Chemsuschem*, **11**, 1600-1605.

Sakamoto, S., Kamimura, N., Tokue, Y., Nakata, M.T., Yamamoto, M., Hu, S., Masai, E., Mitsuda, N. and Kajita, S. (2020) Identification of enzymatic genes with the potential to reduce biomass recalcitrance through lignin manipulation in *Arabidopsis*. *Biotechnol Biofuels*, **13**, 97.

Sakamoto, S., Takata, N., Oshima, Y., Yoshida, K., Taniguchi, T. and Mitsuda, N. (2016) Wood reinforcement of poplar by rice NAC transcription factor. *Sci Rep*, **6**, 19925.

Saleme, M.L.S., Cesarino, I., Vargas, L., Kim, H., Vanholme, R., Goeminne, G., Van Acker, R., Fonseca, F.C.A., Pallidis, A., Voorend, W., Junior, J.N., Padmakshan, D., Van Doorsselaere, J., Ralph, J. and Boerjan, W. (2017) Silencing *CAFFEOYL*

SHIKIMATE ESTERASE affects lignification and improves saccharification in poplar. *Plant Physiol*, **175**, 1040-1057.

Sato, T., Takabe, K. and Fujita, M. (2004) Immunolocalization of phenylalanine ammonia-lyase and cinnamate-4-hydroxylase in differentiating xylem of poplar. *Cr Biol*, **327**, 827-836.

Serrani-Yarce, J.C., Escamilla-Trevino, L., Barros, J., Gallego-Giraldo, L., Pu, Y., Ragauskas, A. and Dixon, R.A. (2021) Targeting hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase for lignin modification in *Brachypodium distachyon*. *Biotechnol Biofuels*, **14**, 50.

Shadle, G., Chen, F., Reddy, M.S.S., Jackson, L., Nakashima, J. and Dixon, R.A. (2007) Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. *Phytochem*, **68**, 2023-2023.

Shi, J., Pattathil, S., Parthasarathi, R., Anderson, N.A., Kim, J.I., Venketachalam, S., Hahn, M.G., Chapple, C., Simmons, B.A. and Singh, S. (2016) Impact of engineered lignin composition on biomass recalcitrance and ionic liquid pretreatment efficiency. *Green Chem*, **18**, 4884-4895.

Steenackers, W., Klima, P., Quareshy, M., Cesarino, I., Kumpf, R.P., Corneillie, S., Araujo, P., Viaene, T., Goeminne, G., Nowack, M.K., Ljung, K., Friml, J., Blakeslee, J.J., Novak, O., Zazimalova, E., Napier, R., Boerjan, W. and Vanholme, B. (2017) *cis*-Cinnamic acid Is a novel, natural auxin efflux inhibitor that promotes lateral root formation. *Plant Physiol*, **173**, 552-565.

Stewart, J.J., Akiyama, T., Chapple, C., Ralph, J. and Mansfield, S.D. (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiol*, **150**, 621-635.

Sykes, R.W., Gjersing, E.L., Foutz, K., Rottmann, W.H., Kuhn, S.A., Foster, C.E., Ziebell, A., Turner, G.B., Decker, S.R., Hinchee, M.A. and Davis, M.F. (2015) Down-regulation of p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H) and cinnamate 4-hydroxylase (C4H) genes in the lignin biosynthetic pathway of *Eucalyptus urophylla* x *E. grandis* leads to improved sugar release. *Biotechnol Biofuels*, **8**, 128.

Takabe, K., Takeuchi, M., Sato, T., Ito, M. and Fujita, M. (2001) Immunocytochemical localization of enzymes involved in lignification of the cell wall. *J Plant Res*, **114**, 509-515.

Tolbert, A., Akinosho, H., Khunsupat, R., Naskar, A.K. and Ragauskas, A.J. (2014) Characterization and analysis of the molecular weight of lignin for biorefining studies. *Biofuel Bioprod Biorefin*, **8**, 836-856.

Vanholme, B., Cesarino, I., Goeminne, G., Kim, H., Marroni, F., Van Acker, R., Vanholme, R., Morreel, K., Ivens, B., Pinosio, S., Morgante, M., Ralph, J., Bastien, C. and Boerjan, W. (2013a) Breeding with rare defective alleles (BRDA): a natural *Populus nigra* HCT mutant with modified lignin as a case study. *New Phytol*, **198**, 765-776.

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y.G., Sundin, L., Goeminne, G., Kim, H., Cross, J., Morreel, K., Araujo, P., Welsh, L., Haustraete, J., McClellan, C., Vanholme, B., Ralph, J., Simpson, G.G., Halpin, C. and Boerjan, W. (2013b) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science*, **341**, 1103-1106.

Vanholme, R., De Meester, B., Ralph, J. and Boerjan, W. (2019) Lignin biosynthesis and its integration into metabolism. *Curr Opin Biotech*, **56**, 230-239.

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

Wagner, A., Ralph, J., Akiyama, T., Flint, H., Phillips, L., Torr, K., Nanayakkara, B. and Te Kiri, L. (2007) Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyltransferase in *Pinus radiata*. *Proc Natl Acad Sci U S A*, **104**, 11856-11861.

Xue, J., Luo, D., Xu, D., Zeng, M., Cui, X., Li, L. and Huang, H. (2015) CCR1, an enzyme required for lignin biosynthesis in Arabidopsis, mediates cell proliferation exit for leaf development. *Plant J*, **83**, 375-387.

Yokoyama, R., de Oliveira, M.V.V., Kleven, B. and Maeda, H.A. (2021) The entry reaction of the plant shikimate pathway is subjected to highly complex metabolite-mediated regulation. *Plant Cell*, **33**, 671-696.

Yoo, C.G., Dumitrache, A., Muchero, W., Natzke, J., Akinosho, H., Li, M., Sikes, R.W., Brown, S.D., Davison, B., Tuskan, G.A., Pu, Y.Q. and Ragauskas, A.J. (2018) Significance of lignin S/G ratio in biomass recalcitrance of *Populus trichocarpa* variants for bioethanol production. *Acs Sustain Chem Eng*, **6**, 2162-2168.

Yoshida, K., Sakamoto, S., Kawai, T., Kobayashi, Y., Sato, K., Ichinose, Y., Yaoi, K., Akiyoshi-Endo, M., Sato, H., Takamizo, T., Ohme-Takagi, M. and Mitsuda, N. (2013) Engineering the *Oryza sativa* cell wall with rice NAC transcription factors regulating secondary wall formation. *Front Plant Sci*, **4**, 383.

Yue, F., Lu, F., Sun, R.C. and Ralph, J. (2012) Syntheses of lignin-derived thioacidolysis monomers and their uses as quantitation standards. *J Agric Food Chem*, **60**, 922-928.

Zhao, Y., Yu, X., Lam, P.Y., Zhang, K., Tobimatsu, Y. and Liu, C.J. (2021) Monolignol acyltransferase for lignin *p*-hydroxybenzoylation in *Populus*. *Nat Plants*, **7**, 1288-1300.

Zhou, X., Ren, S., Lu, M., Zhao, S., Chen, Z., Zhao, R. and Lv, J. (2018) Preliminary study of cell wall structure and its mechanical properties of C3H and HCT RNAi transgenic poplar sapling. *Sci Rep*, **8**, 10508.

Ziebell, A., Gracom, K., Katahira, R., Chen, F., Pu, Y., Ragauskas, A., Dixon, R.A. and Davis, M. (2010) Increase in 4-coumaryl alcohol units during lignification in alfalfa (*Medicago sativa*) alters the extractability and molecular weight of lignin. *J Biol Chem*, **285**, 38961-38968.

Figure legends

Figure 1. Shikimate, general phenylpropanoid, and lignin biosynthetic pathways. HpSK is predicted to transform cytosolic shikimate to shikimate 3-phosphate. SK, shikimate kinase; HCT, hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase; C3'H, *p*-coumarate 3'-hydroxylase; CSE, caffeoyl shikimate esterase; HpSK, SK from *Helicobacter pylori*.

Figure 2. A typical phenotypic representative of transgenic hybrid aspen and their *HpSK* expression examined using RT-PCR. (a) Transgenic lines expressing *HpSK* and the wild type. Plants were grown in pots for 15 weeks in a conditioned culture room. Scale bar = 20 cm. (b) *HpSK* transcripts in stems were detected by RT-PCR. Levels of 18S ribosomal RNA transcripts were used as a control. Nucleotide sequences of the primers are shown in Table S5.

Figure 3. SK activity of wild type and *HpSK* expression lines. The analysis was performed with three biological replicates of each line. The ultrafiltrate (10 kDa cut, 2 mg/mL) of cell extracts were incubated with 0.5 mM shikimate in the presence of 1 mM ATP, 5 mM $MgCl_2$, and 50 mM KCl in 100 mM Tris-HCl (pH 7.5) at 30 °C. Specific activity was calculated from the amount of S3P synthesized for 30 min. Each value represents the mean ± standard deviation.

Figure 4. Partial 2D ^{1}H–^{13}C correlation (HSQC, heteronuclear single-quantum coherence) NMR spectra of wild-type and *HpSK* transgenic plants with major well-separated correlation peaks color-coded to their assigned structures. Note that all spectra contained bisphenol-A carbonate peaks (BPA, gray color) from polycarbonate tubes used for grinding; overlap with the *p*-hydroxyphenyl group $H_{2/6}$ correlation did not allow quantification of H-units (the structure of which is shown). Quantification was performed using the correlation peak volume integration (uncorrected); sidechain units are on the basis **A** + **B** + **C** + **SD** = 100%; aromatics are on the basis **S** + **G** = 100%; *p*B: *p*-hydroxybenzoate.

Figure 5. Expression profiles of lignin biosynthetic genes in stems of transgenic plants. The analysis was performed with three biological replicates of each line. Values indicate the logarithm of fold change (log2FC) of gene expression in each transgenic line compared to the wild type (obtained from biological triplicates). Abbreviated gene names in bold font indicate major genes involved in monolignol biosynthesis. Underlined values represent significant differences compared to the wild type. A bar graph shows expression level of *HpSK* in different transgenic lines, #13, #15, and #21. An asterisk in the graph indicates significant difference from each other according to Student's *t*-test.

Figure 6. Saccharification efficiency of CWR assessed by enzymatic hydrolysis associated with and without alkaline pretreatments. The analysis was performed with five biological replicates of each line. (a) Saccharification efficiency of the wild type and the two transgenic lines with altered

lignin structure. Before enzymatic hydrolysis, the alkaline pretreatment was performed using 62.5 mM NaOH at 120 °C for 1 h. (b) Lignin content in CWR after the alkaline pretreatment by acetyl bromide method. Asterisks indicate significant differences from the wild type using the Student's *t*-test (* $P < 0.05$, ** $P < 0.01$).

Table legends

Table 1. Lignin content and its monomeric composition. Values represent means ± standard error from five independent biological replicates. Asterisks indicate significant differences from the wild type using the Student's *t*-test (*$0.01 < P < 0.05$; **$P < 0.01$). ND: not detected, i.e., signal under the detection limit.

Table 2. Characterized compounds with a significantly different abundance in line #13 and #21, as compared to wild type (WT). The analysis was performed with five biological replicates of each line. Rt: retention time. Fold-change ratio is given as a positive or negative number for compounds with a higher or lower abundance, respectively, in the transgenic lines as compared to the wild type. Values in bold font indicate significant differences compared to the wild type ($P < 0.05$, Student's *t*-test). dis.: discrete, i.e., compounds that remained below detection limit in one of the lines. HCH: hydroxycyclohexenoyl moiety. [1] H(8-O-4)S' is detected as in-source fragment, co-eluting with its molecular ion m/z 373.1278. [2] Compounds detected as in-source fragments, after neutral loss of H_2O. [3] Compounds detected as formic acid adducts. *, shikimate 3-phosphate isomers are characterized based on their exact *m/z* (theoretical m/z 253.012) and retention time. The authentic standard eluted at 1.13 min.

Table 1. Lignin content and its monomeric composition. Values represent means ± standard error from five independent biological replicates. Asterisks indicate significant differences from the wild type using the Student's *t*-test (*0.01 < *P* < 0.05; **P* < 0.01). ND: not detected, i.e., signal under the detection limit.

Line	Acetyl bromide lignin (% CWR)	Thioacidolysis monomers (relative molar %)			H + G + S (µmol/g AcBr lignin)	S/G
		H	G	S		
WT	22.0 ± 0.2	ND	35.7 ± 0.7	64.3 ± 0.7	2226 ± 186	1.80 ± 0.06
#13	21.4 ± 0.5	9.3 ± 5.0**	30.1 ± 0.5**	59.7 ± 1.0*	1184 ± 571**	1.93 ± 0.05**
#21	21.2 ± 0.4	5.8 ± 0.4**	32.5 ± 0.7**	61.6 ± 0.6**	1676 ± 145**	1.89 ± 0.06*
#15	22.1 ± 0.4	ND	35.0 ± 0.7	65.0 ± 0.7	2128 ± 156	1.86 ± 0.06
#22	21.7 ± 0.5	ND	35.4 ± 1.2	64.6 ± 1.2	2258 ± 81	1.83 ± 0.09

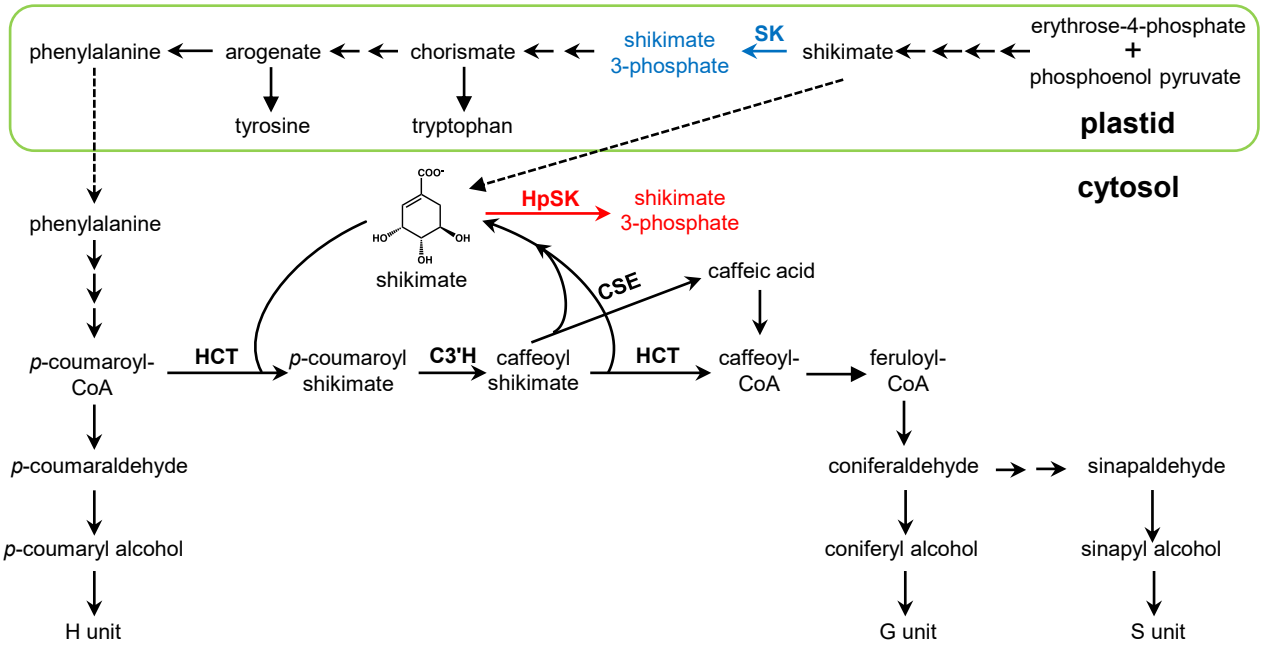
Table 2. Characterized compounds with a significantly different abundance in line #13 and #21, as compared to wild type (WT). The analysis was performed with five biological replicates of each line. Rt: retention time. Fold-change ratio is given as a positive or negative number for compounds with a higher or lower abundance, respectively, in the transgenic lines as compared to the wild type. Values in bold font indicate significant differences compared to the wild type ($P < 0.05$, Student's t-test). dis.: discrete, i.e., compounds that remained below detection limit in one of the lines. HCH: hydroxycyclohexenoyl moiety. [1] H(8-O-4)S' is detected as in-source fragment, co-eluting with its molecular ion m/z 373.1278. [2] Compounds detected as in-source fragments, after neutral loss of H_2O. [3] Compounds detected as formic acid adducts. *, shikimate 3-phosphate isomers are characterized based on their exact *m/z* (theoretical m/z 253.012) and retention time. The authentic standard eluted at 1.13 min.

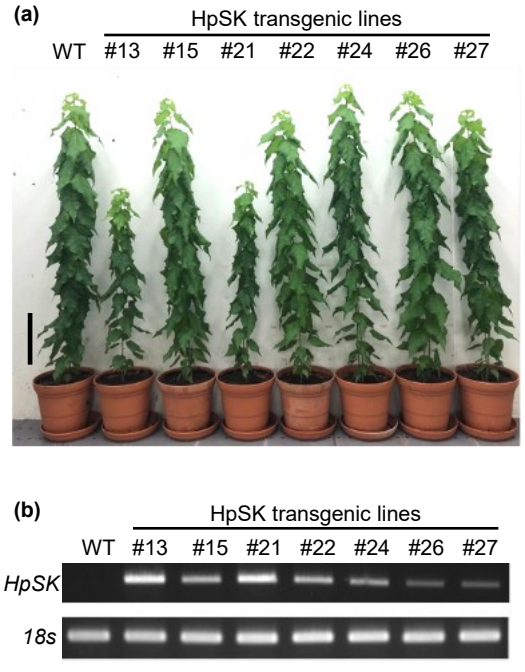
Compounds	**Rt (min)**	***m/z***	**WT control**	***HpSK* #13**	**Fold Change #13/WT**	***P* value**	***HpSK* #21**	**Fold Change #21/WT**	***P* value**
Shikimate derivates									
shikimate-3-phosphate (S3P) 1*	1.04	253.0105	0 ± 0	197 ± 80	**dis.**	**-**	176 ± 83	**dis.**	**-**
shikimate-3-phosphate (S3P) 2*	1.19	253.0103	0 ± 0	98 ± 14	**dis.**	**-**	110 ± 44	**dis.**	**-**
5-enolpyruvylshikimate-3-phosphate (EPSP) 1	2.02	323.0168	76 ± 48	11572 ± 2733	**152**	<0.001	15565 ± 3937	**205**	<0.001
5-enolpyruvylshikimate-3-phosphate (EPSP) 2	2.19	323.0167	203 ± 98	14239 ± 3672	**70**	<0.001	15513 ± 4239	**76**	<0.001
Aromatic amino acid									
phenylalanine	3.36	164.0713	874 ± 246	376 ± 109	**-2.3**	0.002	460 ± 65	**-1.9**	0.011
tyrosine	1.94	180.0662	1442 ± 1639	66 ± 28	-21.8	0.067	236 ± 108	-6.1	0.091
tryptophan	5.04	203.0821	4830 ± 3740	524 ± 237	**-9.2**	0.033	1491 ± 613	-3.2	0.068
Hydroxycinnamoyl shikimate									
caffeoyl shikimate 1	7.94	335.0767	3165 ± 2481	278 ± 139	**-11.4**	0.033	766 ± 339	-4.1	0.059
caffeoyl shikimate 2	8.36	335.0768	13557 ± 13202	833 ± 222	**-16.3**	0.0497	2539 ± 719	-5.3	0.073
caffeoyl shikimate 3	9.82	335.0764	713 ± 729	30 ± 11	-23.8	0.054	120 ± 34	-5.9	0.073

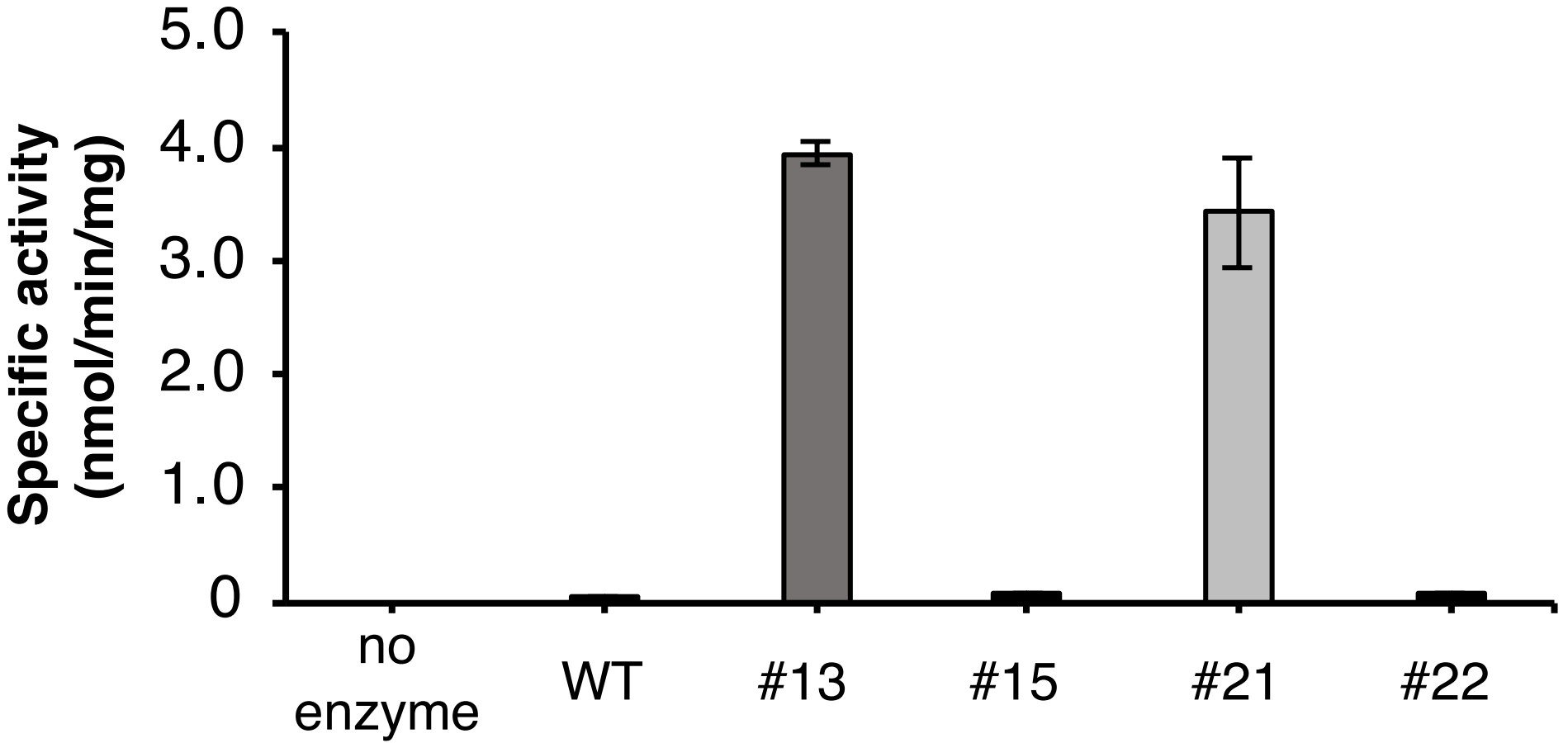
Hydroxycinnamoyl quinate									
caffeoyl quinate (chlorogenic acid)	6.34	353.0869	1701 ± 822	499 ± 97	**-3.4**	0.012	1446 ± 182	-1.2	0.281
feruloyl quinate	6.99	367.1025	354 ± 111	49 ± 14	**-7.2**	0.002	189 ± 16	**-1.9**	0.017
Phenolic acid and hydroxycinnamic acid and derivates									
caffeic acid 3/4-O-hexoside 1	6.12	341.0873	4745 ± 882	11278 ± 1409	**2.4**	<0.001	7995 ± 1871	**1.7**	0.009
caffeic acid 3/4-O-hexoside 2	6.62	341.0872	2079 ± 1959	190 ± 20	**-10.9**	0.048	329 ± 21	-6.3	0.058
caffeoyl hexose 3/4-O-hexoside	5.80	503.1397	607 ± 64	1213 ± 254	**2.0**	0.002	597 ± 137	-1.0	0.434
ferulic acid + hexose sulfate	5.35	435.0594	1952 ± 423	561 ± 295	**-3.5**	0.004	2192 ± 72	1.1	0.139
dihydroferulic acid + hexose	6.93	357.1184	747 ± 104	1487 ± 414	**2.0**	0.003	1141 ± 134	**1.5**	0.003
sinapic acid 4-O-hexoside	6.64	385.1135	5098 ± 1086	17922 ± 1590	**3.5**	<0.001	13455 ± 1780	**2.6**	<0.001
3,4-dihydroxybenzoic acid (protocatechuic acid)	4.42	153.0187	595 ± 370	177 ± 36	**-3.4**	0.031	360 ± 50	-1.7	0120
dihydroxybenzoic acid + pentose	6.14	285.0606	289 ± 208	29 ± 7	**-10.0**	0.023	105 ± 7	-2.8	0.059
vanillic acid 4-O-hexoside	4.18	329.0871	1648 ± 509	541 ± 181	**-3.0**	0.001	1144 ± 132	**-1.4**	0.038
syringoyl hexose	5.37	329.0870	903 ± 271	84 ± 23	**-10.75**	0.001	185 ± 58	**-4.9**	0.002
monolignol, oligolignol and derivatives									
H(8-O-4)H	7.59	315.1226	0 ± 0	97 ± 36	**dis.**	-	75 ± 15	**dis.**	-
H(8-O-4)G'	8.14	343.1179	0 ± 0	242 ± 80	**dis.**	-	222 ± 72	**dis.**	-
H(8-O-4)S' [1]	14.43	207.0658	0 ± 0	458 ± 227	**dis.**	-	204 ± 162	**dis.**	-
H(8-O-4)S(8-5)G'	18.22	551.1915	132 ± 170	4093 ± 3547	**31.0**	0.031	2159 ± 2774	16.4	0.085
G(8-5)G [2]	13.14	339.1231	1425 ± 1064	119 ± 41	**-12.0**	0.026	340 ± 66	**-4.2**	0.042
G(8-5)G [2]	15.02	339.1231	1445 ± 411	215 ± 52	**-6.7**	0.002	245 ± 268	**-5.9**	<0.001
G(8-5)G' [2]	17.36	337.1073	803 ± 811	108 ± 44	-7.4	0.058	91 ± 146	**-8.8**	0.046
G(8-8)G	14.83	357.1337	885 ± 251	147 ± 23	**-6.0**	0.002	248 ± 147	**-3.6**	<0.001
Gox(8-O-4)G	13.77	373.1285	1950 ± 628	507 ± 89	**-3.8**	0.003	635 ± 321	**-3.1**	0.003

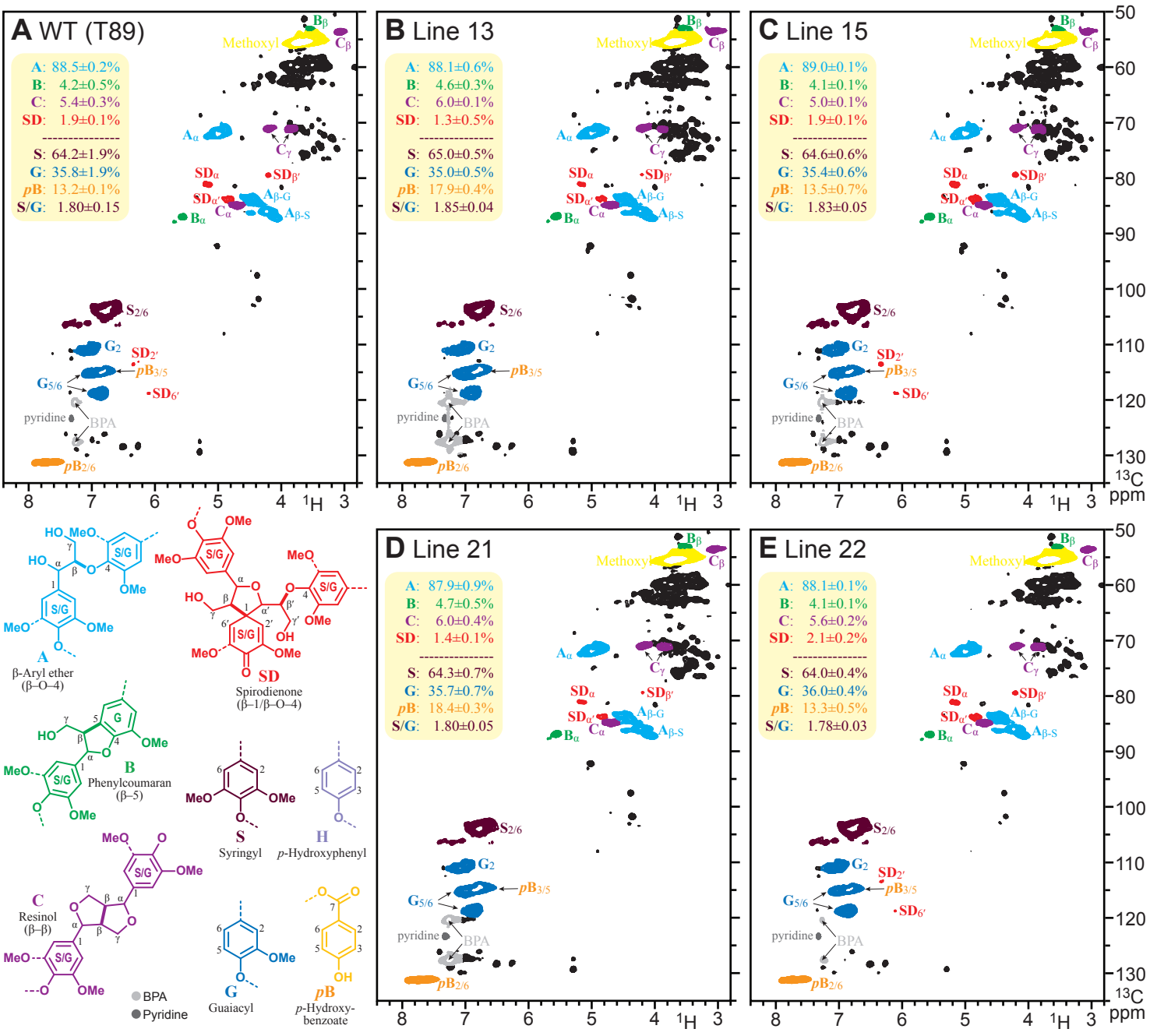
G(8-O-4)G(8-5)G	15.35	553.2069	838 ± 477	159 ± 58	**-5.3**	0.018	136 ± 159	**-6.2**	0.011
G(8-O-4)G(8-O-4)G 1	11.84	571.2175	1504 ± 1283	138 ± 36	**-10.9**	0.039	385 ± 80	-3.9	0.063
G(8-O-4)G(8-O-4)G 2	11.92	571.2177	2142 ± 1723	260 ± 46	**-8.2**	0.035	512 ± 171	-4.2	0.051
G(8-O-4)G(8-O-4)G 3	11.99	571.2176	1152 ± 904	114 ± 23	**-10.1**	0.031	268 ± 77	**-4.3**	0.048
G(8-O-4)G(8-O-4)G 4	12.08	571.2176	2356 ± 1908	224 ± 44	**-10.5**	0.034	524 ± 154	-4.5	0.050
G(8-O-4)G(red8-5)G	15.05	555.2224	1998 ± 963	36 ± 11	**-55.5**	0.005	63 ± 64	**-31.7**	0.005
G 4-O-hexoside(8-O-4)G 1 [3]	8.20	583.2025	1632 ± 1521	168 ± 49	**-9.7**	0.048	311 ± 59	-5.2	0.061
G 4-O-hexoside(8-O-4)G 2 [3]	8.42	583.2026	7661 ± 5066	720 ± 263	**-10.6**	0.0019	1781 ± 162	**-4.3**	0.003
sinapyl alcohol	9.53	209.0815	1495 ± 1125	201 ± 46	**-7.4**	0.029	558 ± 349	-2.7	0.050
G(8-O-4)S^{glycerol}	6.74	439.1596	379 ± 44	45 ± 12	**-8.4**	<0.001	63 ± 29	**-6.0**	<0.001
Gox(8-O-4)S	14.63	403.139	1719 ± 505	226 ± 36	**-7.6**	0.001	469 ± 156	**-3.7**	0.001
G(8-O-4)S(8-5)G 1	17.31	583.218	9155 ± 5309	713 ± 187	**-12.8**	0.001	1829 ± 1455	**-5.0**	0.013
G(8-O-4)S(8-5)G 2	16.56	583.218	35392 ± 10801	4443 ± 925	**-8.0**	0.002	8871 ± 5577	**-4.0**	0.001
S(8-O-4)S(8-5)G	16.34	613.2283	5159 ± 1220	706 ± 95	**-7.3**	<0.001	1422 ± 654	**-3.6**	<0.001
G(8-O-4)S(8-5)G' 1	18.48	581.2023	14153 ± 14828	555 ± 303	-25.5	0.053	1839 ± 1949	-7.7	0.057
G(8-O-4)S(8-5)G' 2	19.19	581.2023	2232 ± 2423	33 ± 29	-67.6	0.055	147 ± 161	-15.2	0.058
G(8-O-4)S(8-8)S 1	18.41	613.2283	6989 ± 2082	630 ± 144	**-11.1**	0.001	1954 ± 1394	**-3.6**	<0.001
G(8-O-4)S(8-8)S 2	19.19	613.2282	1270 ± 600	98 ± 37	**-13.0**	0.005	361 ± 352	**-3.5**	0.008
G(8-O-4)G(8-O-4)S(8-8)S or G(8-O-4)S(8-8)S(8-O-4)G 1	19.57	809.3016	22249 ± 14188	1167 ± 347	**-19.1**	0.014	6265 ± 4411	**-3.6**	0.022
G(8-O-4)G(8-O-4)S(8-8)S or G(8-O-4)S(8-8)S(8-O-4)G 2	20.24	809.3014	9753 ± 9221	571 ± 371	**-17.1**	0.043	2305 ± 2197	**-4.2**	0.065
G(8-O-4)G(8-O-4)S(8-8)S or G(8-O-4)S(8-8)S(8-O-4)G 3	20.87	809.3011	775 ± 922	2 ± 2	-387.5	0.067	80 ± 109	-9.7	0.083
G(8-O-4)G(8-O-4)S(8-5)G	16.75	779.2908	2398 ± 1582	291 ± 98	**-8.2**	0.020	623 ± 528	**-3.8**	0.021
G(8-O-4)S(8-8)G or G(8-O-4)G(8-8)S 1	18.69	583.2177	6206 ± 3123	446 ± 117	**-13.9**	0.007	1029 ± 924	**-6.0**	0.006
G(8-O-4)S(8-8)G or G(8-O-4)G(8-8)S 2	19.46	583.2178	1336 ± 867	35 ± 12	**-38.2**	0.014	155 ± 184	**-8.6**	0.015
S(8-5)G [2]	14.83	369.1337	1235 ± 331	201 ± 37	**-6.1**	0.001	345 ± 200	**-3.6**	<0.001

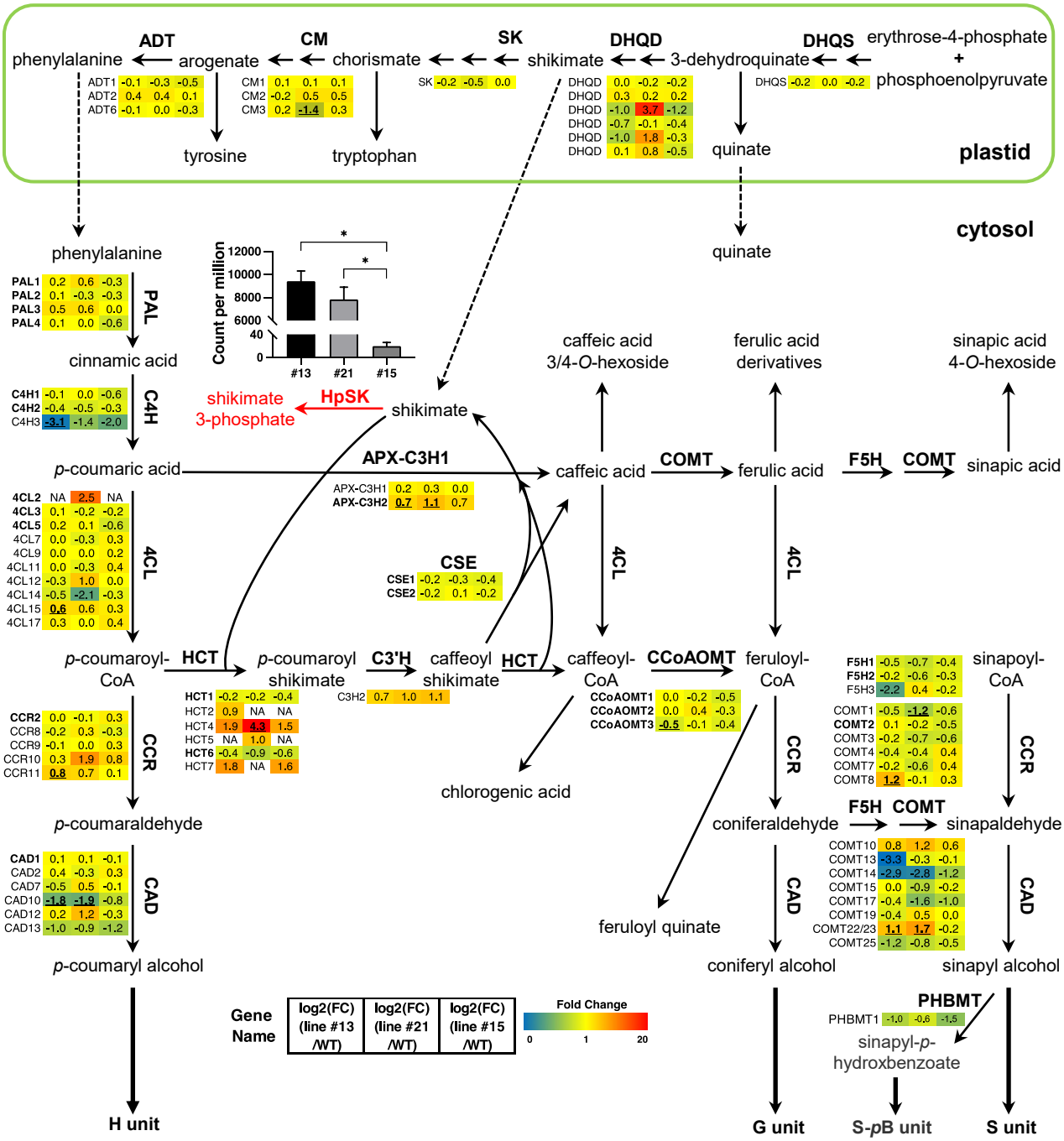
Salicinoids									
trichocarpin	13.91	405.1184	5987 ± 4816	1669 ± 2088	**-3.6**	0.013	2170 ± 638	-2.8	0.082
salicortin	11.30	423.1291	216070 ± 60153	59184 ± 18889	**-3.7**	0.001	170752 ± 20013	-1.3	0.114
tremulacin	20.20	527.1553	60442 ± 51518	7867 ± 2666	**-7.7**	0.038	28855 ± 14890	-2.1	0.155
HCH-salicortin	16.00	561.1606	253903 ± 71127	52860 ± 22112	**-4.8**	<0.001	160338 ± 19852	**-1.6**	0.016
salireposide	13.25	405.1184	8543 ± 6341	2357 ± 1612	**-3.6**	0.023	5189 ± 1902	-1.6	0.181
Other metabolites									
9,12,13-trihydroxy-10(*E*),15(*Z*)-octadecadienoic acid	19.95	327.2171	1150 ± 776	326 ± 228	**-3.5**	0.017	846 ± 376	-1.4	0.276
9,12,13-trihydroxy-10(*E*)-octadecenoic acid	21.52	329.2326	1095 ± 315	240 ± 119	**-4.6**	<0.001	830 ± 438	-1.3	0.220
12-hydroxy jasmonic acid sulfate	7.88	305.0696	29647 ± 22736	6164 ± 1911	**-4.8**	0.045	17168 ± 2403	-1.7	0.142

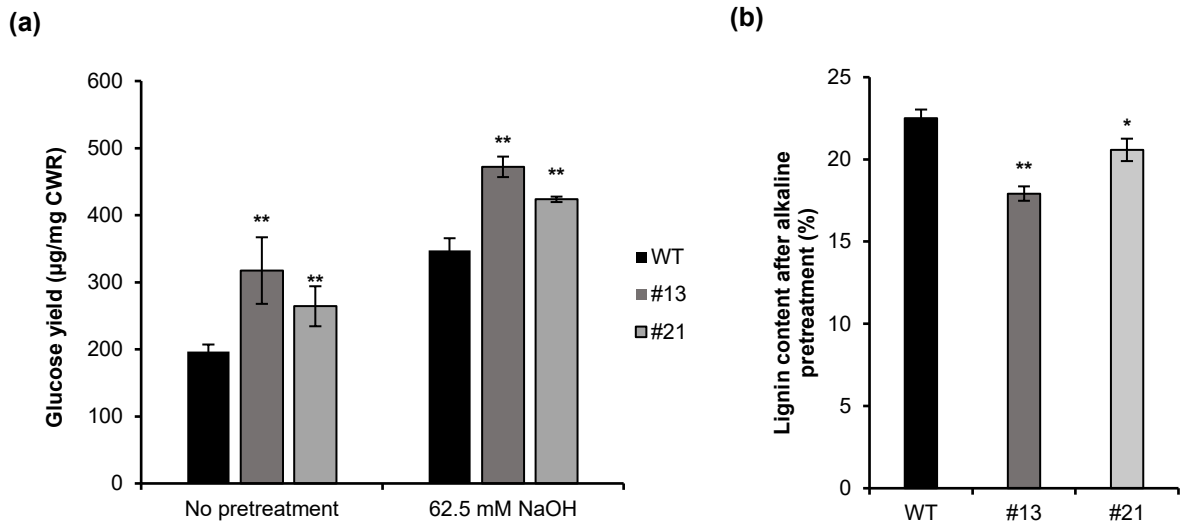

phenylalanine
arogenate
chorismate
shikimate 3-phosphate
SK
shikimate
erythrose-4-phosphate + phosphoenol pyruvate
tyrosine
tryptophan
plastid
cytosol
HpSK
shikimate 3-phosphate
shikimate
phenylalanine
caffeic acid
CSE
p-coumaroyl-CoA
HCT
p-coumaroyl shikimate
C3'H
caffeoyl shikimate
HCT
caffeoyl-CoA
feruloyl-CoA
p-coumaraldehyde
coniferaldehyde
sinapaldehyde
p-coumaryl alcohol
coniferyl alcohol
sinapyl alcohol
H unit
G unit
S unit


(a)
HpSK transgenic lines
WT #13 #15 #21 #22 #24 #26 #27
(b)
HpSK transgenic lines
WT #13 #15 #21 #22 #24 #26 #27
HpSK
18s


Specific activity
(nmol/min/mg)
5.0
4.0
3.0
2.0
1.0
0
no enzyme
WT
#13
#15
#21
#22


A WT (T89)
A: 88.5±0.2%
B: 4.2±0.5%
C: 5.4±0.3%
SD: 1.9±0.1%
S: 64.2±1.9%
G: 35.8±1.9%
pB: 13.2±0.1%
S/G: 1.80±0.15
B Line 13
A: 88.1±0.6%
B: 4.6±0.3%
C: 6.0±0.1%
SD: 1.3±0.5%
S: 65.0±0.5%
G: 35.0±0.5%
pB: 17.9±0.4%
S/G: 1.85±0.04
C Line 15
A: 89.0±0.1%
B: 4.1±0.1%
C: 5.0±0.1%
SD: 1.9±0.1%
S: 64.6±0.6%
G: 35.4±0.6%
pB: 13.5±0.7%
S/G: 1.83±0.05
D Line 21
A: 87.9±0.9%
B: 4.7±0.5%
C: 6.0±0.4%
SD: 1.4±0.1%
S: 64.3±0.7%
G: 35.7±0.7%
pB: 18.4±0.3%
S/G: 1.80±0.05
E Line 22
A: 88.1±0.1%
B: 4.1±0.1%
C: 5.6±0.2%
SD: 2.1±0.2%
S: 64.0±0.4%
G: 36.0±0.4%
pB: 13.3±0.5%
S/G: 1.78±0.03
Methoxyl
Bβ
Cβ
Aα
Cγ
SDα
SDβ′
SDα′
Cα
Aβ-G
Aβ-S
Bα
S2/6
G2
SD2′
pB3/5
G5/6
SD6′
pyridine
BPA
pB2/6
1H
13C
ppm
A
β-Aryl ether
(β–O–4)
SD
Spirodienone
(β–1/β–O–4)
B
Phenylcoumaran
(β–5)
S
Syringyl
H
p-Hydroxyphenyl
C
Resinol
(β–β)
G
Guaiacyl
pB
p-Hydroxy-
benzoate
BPA
Pyridine


plastid
cytosol
ADT
CM
SK
DHQD
DHQS
erythrose-4-phosphate
+
phosphoenolpyruvate
phenylalanine
arogenate
chorismate
shikimate
3-dehydroquinate
quinate
tyrosine
tryptophan
ADT1 -0.1 -0.3 -0.5
ADT2 0.4 0.4 0.1
ADT6 -0.1 0.0 -0.3
CM1 0.1 0.1 0.1
CM2 -0.2 0.5 0.5
CM3 0.2 -1.4 0.3
SK -0.2 -0.5 0.0
DHQD 0.0 -0.2 -0.2
DHQD 0.3 0.2 0.2
DHQD -1.0 3.7 -1.2
DHQD -0.7 -0.1 -0.4
DHQD -1.0 1.8 -0.3
DHQD 0.1 0.8 -0.5
DHQS -0.2 0.0 -0.2
Count per million
12000
10000
8000
6000
40
0
#13
#21
#15
*
*
PAL
PAL1 0.2 0.6 -0.3
PAL2 0.1 -0.3 -0.3
PAL3 0.5 0.6 0.0
PAL4 0.1 0.0 -0.6
cinnamic acid
C4H
C4H1 -0.1 0.0 -0.6
C4H2 -0.4 -0.5 -0.3
C4H3 -3.1 -1.4 -2.0
shikimate 3-phosphate
HpSK
shikimate
caffeic acid 3/4-O-hexoside
ferulic acid derivatives
sinapic acid 4-O-hexoside
p-coumaric acid
APX-C3H1
APX-C3H1 0.2 0.3 0.0
APX-C3H2 0.7 1.1 0.7
caffeic acid
COMT
ferulic acid
F5H
COMT
sinapic acid
4CL
4CL2 NA 2.5 NA
4CL3 0.1 -0.2 -0.2
4CL5 0.2 0.1 -0.6
4CL7 0.0 -0.3 0.3
4CL9 0.0 0.0 0.2
4CL11 0.0 -0.3 0.4
4CL12 -0.3 1.0 0.0
4CL14 -0.5 -2.1 -0.3
4CL15 0.6 0.6 0.3
4CL17 0.3 0.0 0.4
CSE
CSE1 -0.2 -0.3 -0.4
CSE2 -0.2 0.1 -0.2
4CL
4CL
p-coumaroyl-CoA
HCT
p-coumaroyl shikimate
C3'H
caffeoyl shikimate
HCT
caffeoyl-CoA
CCoAOMT
feruloyl-CoA
sinapoyl-CoA
HCT1 -0.2 -0.2 -0.4
HCT2 0.9 NA NA
HCT4 1.9 4.3 1.5
HCT5 NA 1.0 NA
HCT6 -0.4 -0.9 -0.6
HCT7 1.8 NA 1.6
C3H2 0.7 1.0 1.1
CCoAOMT1 0.0 -0.2 -0.5
CCoAOMT2 0.0 0.4 -0.3
CCoAOMT3 -0.5 -0.1 -0.4
F5H1 -0.5 -0.7 -0.4
F5H2 -0.2 -0.6 -0.3
F5H3 -2.2 0.4 -0.2
COMT1 -0.5 -1.2 -0.6
COMT2 0.1 -0.2 -0.5
COMT3 -0.2 -0.7 -0.6
COMT4 -0.4 -0.4 0.4
COMT7 -0.2 -0.6 0.4
COMT8 1.2 -0.1 0.3
CCR
CCR2 0.0 -0.1 0.3
CCR8 -0.2 0.3 -0.3
CCR9 -0.1 0.0 0.3
CCR10 0.3 1.9 0.8
CCR11 0.8 0.7 0.1
chlorogenic acid
CCR
CCR
p-coumaraldehyde
coniferaldehyde
F5H
COMT
sinapaldehyde
COMT10 0.8 1.2 0.6
COMT13 -3.3 -0.3 -0.1
COMT14 -2.9 -2.8 -1.2
COMT15 0.0 -0.9 -0.2
COMT17 -0.4 -1.6 -1.0
COMT19 -0.4 0.5 0.0
COMT22/23 1.1 1.7 -0.2
COMT25 -1.2 -0.8 -0.5
CAD
CAD1 0.1 0.1 -0.1
CAD2 0.4 -0.3 0.3
CAD7 -0.5 0.5 -0.1
CAD10 -1.8 -1.9 -0.8
CAD12 0.2 1.2 -0.3
CAD13 -1.0 -0.9 -1.2
feruloyl quinate
CAD
CAD
p-coumaryl alcohol
coniferyl alcohol
sinapyl alcohol
PHBMT
PHBMT1 -1.0 -0.6 -1.5
sinapyl-p-hydroxbenzoate
Gene Name
log2(FC) (line #13 /WT)
log2(FC) (line #21 /WT)
log2(FC) (line #15 /WT)
Fold Change
0
1
20
H unit
G unit
S-pB unit
S unit


(a)
Glucose yield (µg/mg CWR)
600
500
400
300
200
100
0
**
**
**
**
No pretreatment
62.5 mM NaOH
WT
#13
#21
(b)
Lignin content after alkaline pretreatment (%)
25
20
15
10
5
0
**
*
WT
#13
#21