Plant Mol Biol (2014) 84:565–576
DOI 10.1007/s11103-013-0152-y

Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (*Panicum virgatum* L.)

Luis L. Escamilla-Treviño · Hui Shen · Timothy Hernandez · Yanbin Yin · Ying Xu · Richard A. Dixon

Received: 13 September 2013 / Accepted: 26 October 2013 / Published online: 5 November 2013

Abstract Studying lignin biosynthesis in *Panicum virgatum* (switchgrass) has provided a basis for generating plants with reduced lignin content and increased saccharification efficiency. Chlorogenic acid (CGA, caffeoyl quinate) is the major soluble phenolic compound in switchgrass, and the lignin and CGA biosynthetic pathways potentially share intermediates and enzymes. The enzyme hydroxycinnamoyl-CoA: quinate hydroxycinnamoyltransferase (HQT) is responsible for CGA biosynthesis in tobacco, tomato and globe artichoke, but there are no close orthologs of HQT in switchgrass or in other monocotyledonous plants with complete genome sequences. We examined available transcriptomic databases for genes encoding enzymes potentially involved in CGA biosynthesis in switchgrass. The protein products of two hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) genes (*PvHCT1a* and *PvHCT2a)*, closely related to lignin pathway HCTs from other species, were characterized biochemically and exhibited the expected HCT activity, preferring shikimic acid as acyl acceptor. We also characterized two switchgrass coumaroyl shikimate 3′-hydroxylase (C3′H) enzymes (PvC3′H1 and PvC3′H2); both of these cytochrome P450s had the capacity to hydroxylate 4-coumaroyl shikimate or 4-coumaroyl quinate to generate caffeoyl shikimate or CGA. Another switchgrass hydroxycinnamoyl transferase, PvHCT-Like1, is phylogenetically distant from HCTs or HQTs, but exhibits HQT activity, preferring quinic acid as acyl acceptor, and could therefore function in CGA biosynthesis. The biochemical features of the recombinant enzymes, the presence of the corresponding activities in plant protein extracts, and the expression patterns of the corresponding genes, suggest preferred routes to CGA in switchgrass.

Keywords Phenylpropanoid pathway · Lignin · Flavonoids · Chlorogenic acid

Accessions numbers *PvHCT1a*: AB723827, *PvHCT2a*: KC696573, *PvHCT-Like1*: JX845714, *PvC3′H1*: AB723823, *PvC3′H2*: AB723824.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-013-0152-y) contains supplementary material, which is available to authorized users.

L. L. Escamilla-Treviño · H. Shen · T. Hernandez · R. A. Dixon
Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA

L. L. Escamilla-Treviño · H. Shen · T. Hernandez · Y. Yin · Y. Xu · R. A. Dixon
Oak Ridge National Laboratory, BioEnergy Science Center (BESC), Oak Ridge, TN 37831, USA

Present Address:
L. L. Escamilla-Treviño · H. Shen · R. A. Dixon (✉)
Department of Biological Sciences, University of North Texas, 1155 Union Circle #305220, Denton, TX 6203-5017, USA
e-mail: Richard.Dixon@unt.edu

Y. Yin · Y. Xu
Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

Present Address:
Y. Yin
Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115-2857, USA

Introduction

Panicum virgatum (switchgrass) is a monocotyledonous C4 perennial plant native to North America, ranging from the south of Canada to the north of México, and a significant constituent of the North American tallgrass prairie. Since the mid-1980′s it has been investigated as a bioenergy crop, but research on switchgrass has exponentially increased in the US due to the government's investment in large-scale research programs aimed at making cellulosic biofuels cost-competitive. Switchgrass was selected as one of the most important perennial bioenergy feedstocks for the US because of its moderate to high yields on marginal lands, its efficient use of water and nitrogen, its low input requirements, its wide adaptation, and its environmental benefits related to soil conservation and low net greenhouse gas emissions (McLaughlin and Adams Kszos 2005; Schmer et al. 2008; Yuan et al. 2008; Wullschleger et al. 2010).

The resistance of cell walls to digestion to fermentable sugars, called recalcitrance, is the main obstacle in making cellulosic biofuels cost competitive. Lignin limits the accessibility of enzymes for hydrolysis of cellulose and hemicelluloses in secondary cell walls, and is believed to be a major component of recalcitrance. A number of studies in different plant species have confirmed the link between reduced lignin content and improved sugar release (Grabber 2005; Davison et al. 2006; Ralph et al. 2006; Talukder 2006; van der Rest et al. 2006; Chen and Dixon 2007; Jackson et al. 2008). Therefore studying lignin biosynthesis in switchgrass has been a logical approach to developing plants with reduced lignin content and increased saccharification efficiency (Escamilla-Trevino et al. 2010; Fu et al. 2011; Shen et al. 2011; Xu et al. 2011).

Chlorogenic acid (CGA) is an ester formed between caffeic acid and (L)-quinic acid, and is the most abundant soluble phenolic compound in Solanaceous species like tomato, potato, and eggplant. It also accumulates to significant concentrations in apples, artichoke, coffee, pears, and plumbs (Niggeweg et al. 2004; Comino et al. 2009; Sonnante et al. 2010). CGA is also the major soluble phenolic compound in switchgrass (Shen et al. 2009). CGA is believed to function as an antioxidant, and be involved in resistance to insects, in defense against fungal pathogens, and in enzymatic browning of vegetables and fruits. CGA also acts as an antioxidant in the human diet, and protects against degenerative age-related diseases. Some of the dietary CGA is absorbed by the small intestine, but most of it is hydrolyzed in the large intestine by esterases of the microflora to release caffeic acid which is absorbed easily and has similar antioxidant capacity (Niggeweg et al. 2004; Comino et al. 2009; Sonnante et al. 2010).

The biosynthetic pathways to lignin and CGA potentially share common intermediates and enzymes (Supplemental Fig. S1); because of the high concentration of CGA in switchgrass, understanding CGA biosynthesis in this species could therefore be helpful for optimal engineering of the lignin pathway. However, the pathway(s) to CGA in monocots has yet to be elucidated. Studies in the dicots *Solanum lycopersicum* (tomato), *Nicotiana tabacum* (tobacco) and *Cynara cardunculus* (globe artichoke) have suggested three different routes for CGA biosynthesis in plants. In the first pathway, the enzyme hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT) catalyzes the formation of CGA from caffeoyl CoA and quinic acid (Supplemental Fig. S1, route 1). To produce the caffeoyl CoA substrate, 4-coumaroyl CoA is converted to 4-coumaroyl shikimate by hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyltransferase (HCT), this shikimate ester is hydroxylated by the enzyme 4-coumaroyl shikimate 3′-hydroxylase (C3′H), and the caffeoyl shikimate generated is converted to caffeoyl CoA by an HCT enzyme acting in the reverse direction. The second suggested route proceeds by way of synthesis of 4-coumaroyl quinate by HCT or HQT enzymes, followed by hydroxylation of the coumaroyl moiety by C3′H (Supplemental Fig. S1, route 2). In a third potential route, caffeoyl glucoside is the activated intermediate (Supplemental Fig. S1 route 3) (Villegas and Kojima 1986).

The main functions of HCT and C3′H appear to be in the early steps of lignin biosynthesis. Plants down-regulated in the expression of these enzymes have considerable reductions in lignin content and altered lignin composition characterized by increased 4-hydroxyphenyl (H) lignin units and severe reductions in guaiacyl (G) and syringyl (S) units. These changes are associated with growth defects resulting in plants that are smaller to severely dwarf. All these studies have been done in dicotyledonous plants (Hoffmann et al. 2004; Reddy et al. 2005; Chen et al. 2006). HQT has been shown directly to be involved in CGA biosynthesis in tobacco, tomato and globe artichoke; it's down-regulation using transient assays in *Nicotiana benthamiana* or stable transformation in tomato leads to a considerable reduction in CGA content (up to 98 % reduction in tomato). Conversely, transient HQT over-expression in *N. benthamiana* leads to a substantial increase in CGA levels, as does over-expression through stable transformation in tomato (Niggeweg et al. 2004; Sonnante et al. 2010).

Although switchgrass accumulates high levels of CGA (Shen et al. 2009), we were unable to find any gene likely to encode an HQT enzyme in switchgrass databases or in the complete genomes from four other monocotyledonous plants (*Zea mays*, *Sorghum bicolor*, *Oryza sativa* and *Brachypodium distachyon*). We therefore concentrated on a detailed characterization of HCT enzymes from switchgrass; PvHCT1a, PvHCT2a and PvHCT-Like1. The latter enzyme is phylogenetically far from HCTs or HQTs, but

exhibits HQT activity and could, along with the HCTs, be involved in CGA biosynthesis. We also characterized two switchgrass C3′H enzymes (PvC3′H1 and PvC3′H2); both were able to hydroxylate 4-coumaroyl shikimate or 4-coumaroyl quinate to generate caffeoyl shikimate or CGA. Possible parallel pathways for CGA and lignin biosynthesis in switchgrass are discussed.

Results

Isolation and sequence analysis of switchgrass HCT and C3′H cDNAs

PvHCT1a, *PvHCT2a*, *PvC3′H1a* and *PvC3′H2* cDNA sequences were identified in public databases after BLAST queries with sequences that encode previously characterized HCT or C3′H enzymes from other species including *Arabidopsis thaliana*, *Medicago truncatula*, *Nicotiana tabaccum*, *Coffea canephora*, and *Cynara cardunculus*. *PvHCT-Like1* was discovered from transcriptome analysis of lignifying switchgrass cell suspensions and elongating internodes (Shen et al. 2013). Full length cDNA sequences were used for primer design. A phylogenetic analysis was then performed using the encoded protein sequences of *PvHCT1a*, *PvHCT2a* and *PvHCT-Like1*, previously characterized HCT and/or HQT protein sequences from the dicotyledonous plants *A. thaliana* (Hoffmann et al. 2004), *Nicotiana tabaccum* (Hoffmann et al. 2003), *C. canephora* (Lepelley et al. 2007), *Solanum lycopersicum* (Niggeweg et al. 2004) and *C. cardunculus* (Comino et al. 2007; Comino et al. 2009), and closely related protein sequences from monocots (*Z. mays*, *S. bicolor*, *O. sativa* and *B. distachyon*). *PvHCT1a* encoded the most similar protein to previously characterized HCTs from dicotyledonous plants, with between 65 and 67 % amino acid similarity, with PvHCT2a also closely related (between 61 and 63 % amino acid similarity). Both PvHCT1a and PvHCT2a have higher similarities to functionally characterized HCT enzymes than to HQT enzymes (Supplemental Table S1).

Members of the acyltranferase family contain the signature motifs HXXXDG and DFGWG (Niggeweg et al. 2004). The HXXXDG motif in HCTs from *A. thaliana*, *N. tabaccum* and *C. cardunculus* is HHAADG whereas this is replaced by HTLSDG in the HQTs from *N. tabaccum* and *C. cardunculus*. This signature motif in PvHCT1a and PvHCT2a is HHVADG, which is more similar to that of HCTs than HQTs, with only a single conserved amino acid change (Ala to Val) from the HCT signature.

Orthologous genes to *PvHCT1a* and *PvHCT2a* were identified in the genomes of the monocots *Z. mays*, *S. bicolor*, *O. sativa*, and *B. distachyon* by phylogenetic analysis of encoded proteins, and they all fall in the same clade

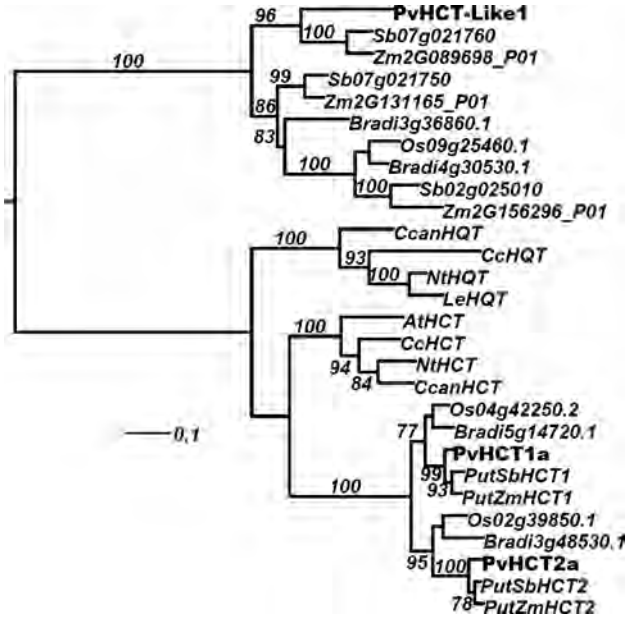

Fig. 1 Phylogenetic tree of *PvHCT1a*, *PvHCT2a* and *PvHCT-Like1*, along with previously characterized HCT and HQT protein sequences from dicotyledonous species; C*offea canephora (Ccan), Cynara cardunculus (Cc), Nicotiana tabaccum* (Nt), *Solanum lycopersicum (Sly) and Arabidopsis thaliana (At)*, and closely related protein sequences from the monocots *Zea mayz (Zm), Sorghum bicolor (Sb), Oryza sativa (Os)* and *Brachypodium distachyon (Bradi)*. The approximate maximum likelihood tree was built using FastTree 2 (see Materials and Methods). Protein IDs are: ZmHCT-1put, NM_001152399; ZmHCT2put, NM_001139418; SbHCT1put, XM_00244670; SbHCT2put, XM_002452390; AtHCT, ABH04595; NtHCT, CAD47830; CcHCT, AAZ80046; NtHQT, CAE46932; CcHQT, ABK79690; LeHQT, CAE46933; CcanHQT, ABO77956; CcanHCT, ABO47805

as PvHCT1a and PvHCT2a (Fig. 1). All the encoded proteins in these two clades are the closest found in these monocotyledonous plants to previously characterized HCTs or HQTs from dicots, and all are more closely related to HCTs than to HQTs (Fig. 1). PvHCT-Like1 has low similarity to PvHCT1a or PvHCT2a and functionally characterized HCTs or HQTs (between 34 to 38 %) (Supplemental Table S1), hence the designation HCT-like. *Z. mays* and *S. bicolor* contain putative orthologs of *PvHCT-Like1* (Fig. 1).

We also performed a phylogenetic analysis of the encoded protein sequences of PvC3′H1, PvC3′H2, its putative orthologous proteins from *S. bicolor* and *Z. mays*, and previously characterized C3′Hs from *C. canephora* (Mahesh et al. 2007), *A. thaliana* (Schoch et al. 2001), *C. cardunculus* (Moglia et al. 2009) and *Triticum aestivum* (Morant et al. 2007) (Supplemental Fig. S2). The two switchgrass C3′Hs exhibit 78 % similarity. PvC3′H1 has 70, 67 and 68 % similarity to the previously characterized C3′Hs from *A. thaliana*, *C. canephora*, and *C. cardunculus* respectively, whereas PvC3′H2 has 67, 64 and 64 %

Table 1 Kinetic parameters for recombinant PvHCT1a, PvHCT2a and PvHCT-Like1

Varying substrate	Saturating substrate	Km (μM^{-1})	Kcat (s^{-1})	Kcat/Km ($s^{-1}\,\mu M^{-1}$)
PvHCT1a				
4-Coumaroyl CoA	Shikimic acid	19.1	60	3.13
4-Coumaroyl CoA	Quinic acid	23.3	8.2	0.35
Caffeoyl CoA	Shikimic acid	3.4	6.1	1.78
Caffeoyl CoA	Quinic acid	5.5	0.054	0.01
Caffeoyl shikimate	CoA	60.6	7.5	0.123
Caffeoyl quinate (CGA)	CoA	No conversion		
PvHCT2a				
4-Coumaroyl CoA	Shikimic acid	5.13	10.3	2
4-Coumaroyl CoA	Quinic acid	4.82	1.13	0.234
Caffeoyl CoA	Shikimic acid	Not determined*	Not determined	Not determined
Caffeoyl CoA	Quinic acid	4.6	0.031	0.007
Caffeoyl shikimate	CoA	56.7	5.2	0.09
Caffeoyl quinate (CGA)	CoA	No conversion		
PvHCT-Like1				
4-Coumaroyl CoA	Shikimic acid	No conversion		
4-coumaroyl CoA	Quinic acid	4.92	3.64	0.74
Caffeoyl CoA	Shikimic acid	No conversion		
Caffeoyl CoA	Quinic acid	4.98	0.77	0.154
Caffeoyl shikimate	CoA	No conversion		
Caffeoyl quinate (CGA)	CoA	No conversion		

* indicates parameters not determined because of strange kinetic behavior

similarity to the same C3′Hs. The closest protein sequences to PvC3′H1 are SbCYP98A1 and ZmCYP98A29, with 96 and 95 % similarity, respectively, while the closest protein sequences to PvC3′H2 are putZmC3′H2 and putSbC3′H2, both with 88 % identity (Supplemental Table S2).

Biochemical characterization of recombinant switchgrass HCT and C3′H enzymes

Recombinant PvHCT1a, PvHCT2a and PvHCT-Like1 were expressed in *E. coli* as His-tagged fusion proteins. They were purified to single bands on SDS-PAGE analysis (Supplemental Fig. S4), migrating to just above the 50 kD ladder marker, in agreement with the calculated sizes of the recombinant enzymes plus the 6-His motif. The recombinant enzymes were assayed with 4-coumaroyl-CoA or caffeoyl-CoA as acyl donors and shikimic acid or quinic acid as acceptors. The reverse reactions were also tested using caffeoyl shikimate/quinate (CGA) and CoA. Kinetic parameters are shown in Table 1.

Both PvHCT1a and PvHCT2a have preference towards shikimic acid as acyl acceptor, the enzyme efficiency (Kcat/Km) with quinic acid being approximately one order of magnitude less for both recombinant enzymes. In contrast, PvHCT-Like1 prefers quinic acid as acyl acceptor with either 4-coumaroyl CoA or caffeoyl CoA as acyl donor. PvHCT-Like1 has preference for 4-coumaroyl CoA over caffeoyl CoA, with an approximately five-fold higher Kcat/Km ratio (0.74 vs 0.15 $s^{-1}\ \mu M^{-1}$). Neither of the three enzymes is able to perform the reverse reaction to generate caffeoyl CoA from CGA, but PvHCT1a and PvHCT2a can generate caffeoyl CoA from caffeoyl shikimate and CoA although quite inefficiently (Kcat/Km = 0.123 or 0.09 $s^{-1}\mu M^{-1}$ respectively). The most efficient reaction occurred with PvHCT1a, 4-coumaroyl CoA and shikimic acid as substrates, with the Kcat/Km value (3.13 $s^{-1}\mu M^{-1}$) being about twofold higher than the value obtained with caffeoyl CoA (Table 1).

Recombinant PvC3′H1 and PvC3′H2 were expressed in the yeast *Saccharomyces cerevisiae* strain WAT11, which contains an *A. thaliana* P450 reductase that can be co-induced with the recombinant P450 and assists in electron transfer to the P450 as well as helping ensure correct enzyme folding (Pompon et al. 1996). Isolated microsomes from induced yeast containing either the PvC3′H1 or PvC3′H2 cDNA had the ability to convert 4-coumaroyl shikimate or 4-coumaroyl quinate to their respective products caffeoyl shikimate or caffeoyl quinate (CGA) (Fig. 2).

Enzyme activities in switchgrass stem extracts

To compare the enzyme activities observed with recombinant enzymes with the reactions that may occur *in planta*,

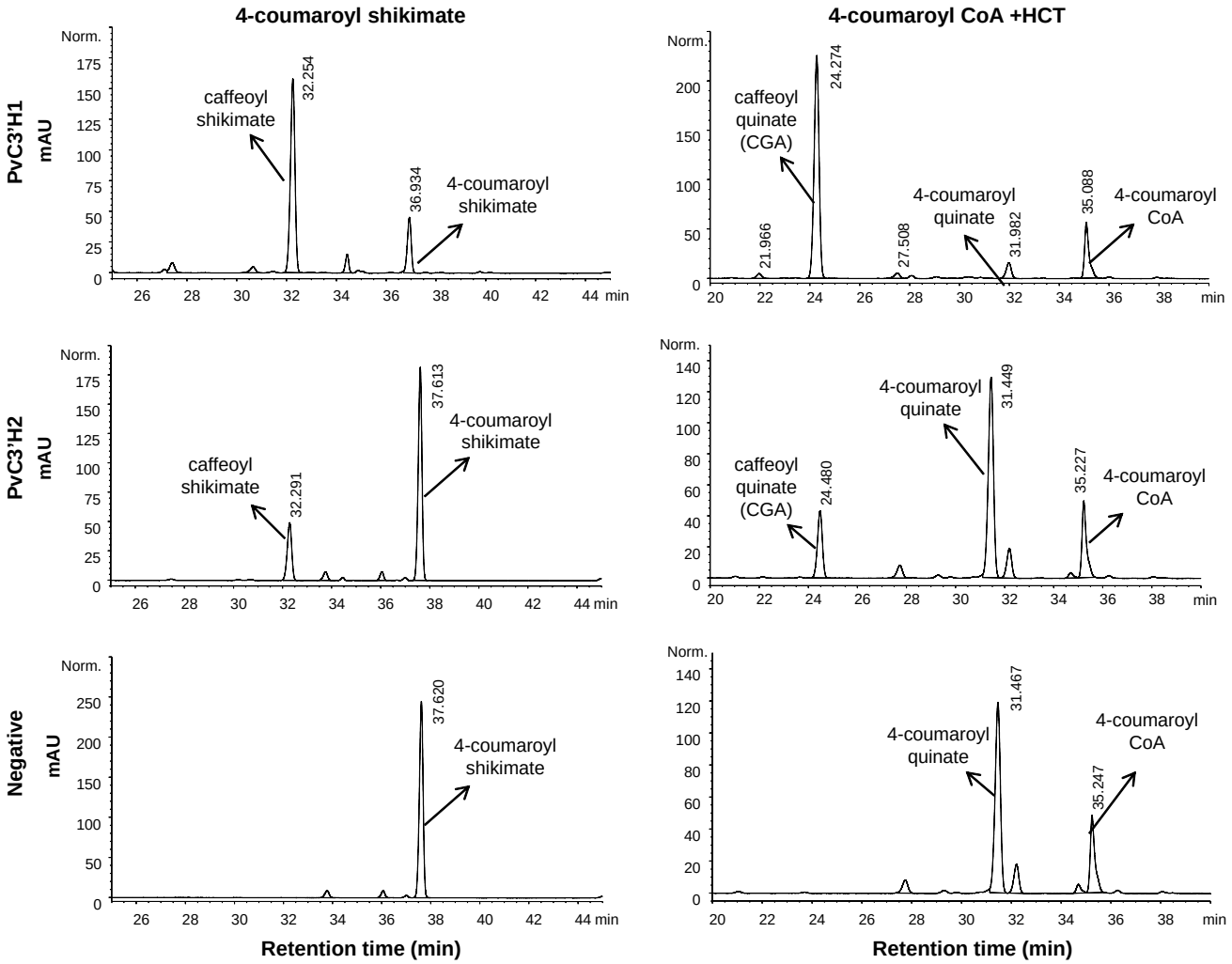

Fig. 2 Chromatograms of products of recombinant C3′H activity assays. Right side shows the reaction using 4-coumaroyl shikimate and left side shows the reaction using 4-coumaroyl CoA, quinic acid and HCT enzyme (to generate the substrate 4-coumaroyl quinate). Microsomes used in the assays were from recombinant yeast expressing either PvC3′H1 (*top*), PvC3′H2 (*middle*), or empty vector (*bottom*)

protein extracts were prepared from switchgrass Alamo ST2 plants at the E2 developmental stage (Moore et al. 1991). Samples were taken from the stem above the first node and leaves. As with the recombinant enzymes, the protein extracts were assayed with 4-coumaroyl-CoA or caffeoyl-CoA as acyl donors and shikimic acid or quinic acid as acyl acceptors, and the reverse reactions were also measured.

The formation of 4-coumaroyl shikimate from 4-coumaroyl-CoA and shikimic acid was the most efficient reaction carried out by the plant extracts, with almost total conversion occurring with both stem and leaf extracts. In contrast, formation of 4-coumaroyl quinate was only just detectable, with most of the 4-coumaroyl-CoA substrate remaining unconverted (Fig. 3). The reverse HCT reaction for formation of caffeoyl CoA from caffeoyl shikimate plus CoA could not be determined in crude extracts. Instead, caffeic acid was the main product formed from caffeoyl shikimate, and this conversion occurred in the presence or absence of CoA (Fig. 4). Recently, a caffeoyl shikimate esterase (CSE, encoded by At1g52760) was reported in Arabidopsis, and shown to play an important role in lignin biosynthesis. CSE, together with 4-coumarate:CoA ligase (4CL) active with caffeic acid, bypasses the second HCT reaction, hence rewriting this part of the pathway (Vanholme et al. 2013). The reverse HCT reaction for formation of caffeoyl CoA from CGA and CoA could also not be shown in the crude extracts, but CGA was not hydrolysed to caffeic acid; instead, an unidentified peak appeared with earlier retention time than CGA (Fig 4). No products were

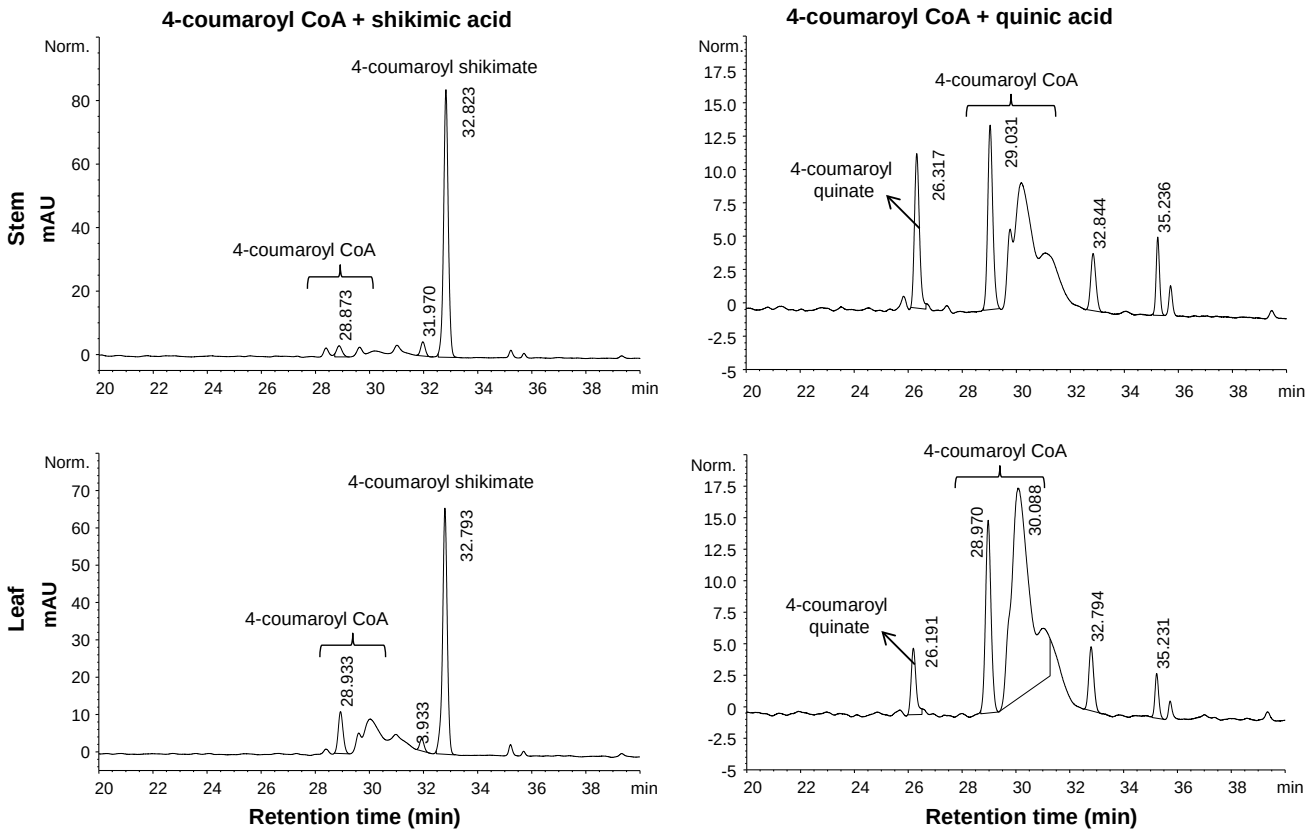

Fig. 3 HPLC chromatograms showing products of HCT activity assays with switchgrass protein extracts from stem (*top panels*) or leaf (*bottom panels*). The substrate was 4-coumaroyl CoA plus either shikimic acid (*left panels*) or quinic acid (*right panels*)

formed when the reaction was repeated in the absence of CoA.

CGA content in switchgrass tissues

As a basis for understanding possible relationships between gene expression and CGA levels, we performed a detailed analysis to determine the CGA content in various plant tissues from E4 stage switchgrass plants of line Alamo ST2. We analyzed leaves 1 and 3 (L1, L3), and internodes 2, 3 and 4 (I2, I3, I4). Under this nomenclature I4 is younger than I2 or I3 (Shen et al. 2009). Since it is known that the top part of the young internodes is more lignified than the bottom part (Shen et al. 2009), we divided the internodes in half so we could separately analyze the bottom (bot) and top parts. The difference in CGA content between the various tissues was quite large; I4-bot (highest content) had about 14 times more CGA than I2-bot (lowest content) (Fig. 5). I4-bot, L1 and L2 had the highest CGA contents, suggesting that CGA accumulates primarily in young and less lignified tissues. The difference of CGA concentration within one internode was also surprisingly high; for example, I4-bot had about sixfold higher CGA level than I4-top (Fig. 5).

Developmental expression of switchgrass *HCT* and *C3′H* Genes

A switchgrass gene expression atlas (PviGEA) was recently developed (Zhang et al. 2013), and transcript expression data for close to 80,000 genes in up to 24 different tissues are publicly available on its web server (http://switchgrassgenomics.noble.org/). We extracted the transcript data from the PviGEA web server for the five genes targeted in this study in tissues with information about lignin and CGA content of E4 stage plants: leaf blade (LFB), leaf sheath (LSH), the middle part of internode 3 (I3-mid), and the vascular bundles of I3-mid (I3-midVB), top, middle and bottom parts of internode 4 (I4-top, I4-mid, I4-bot). The transcript data plotted in a bar graph (Fig. 6) show that *PvHCT1a*, *PvHCT2a*, and *PvC3′H1* have higher transcript levels in I3-mid, I3-midVB, I4-top, and I4-mid (more lignified tissue) than LFB, LSH and I4-bot (less lignified), whereas *PvHCT-Like1* and *PvC3′H2* show higher transcript levels in tissues that have a higher content of CGA and lower content of lignin, such as LFB, LSH and I4-bot (with the exception of *PvC3′H2* in LFB). It was also observed that the transcript levels of the genes varied significantly within

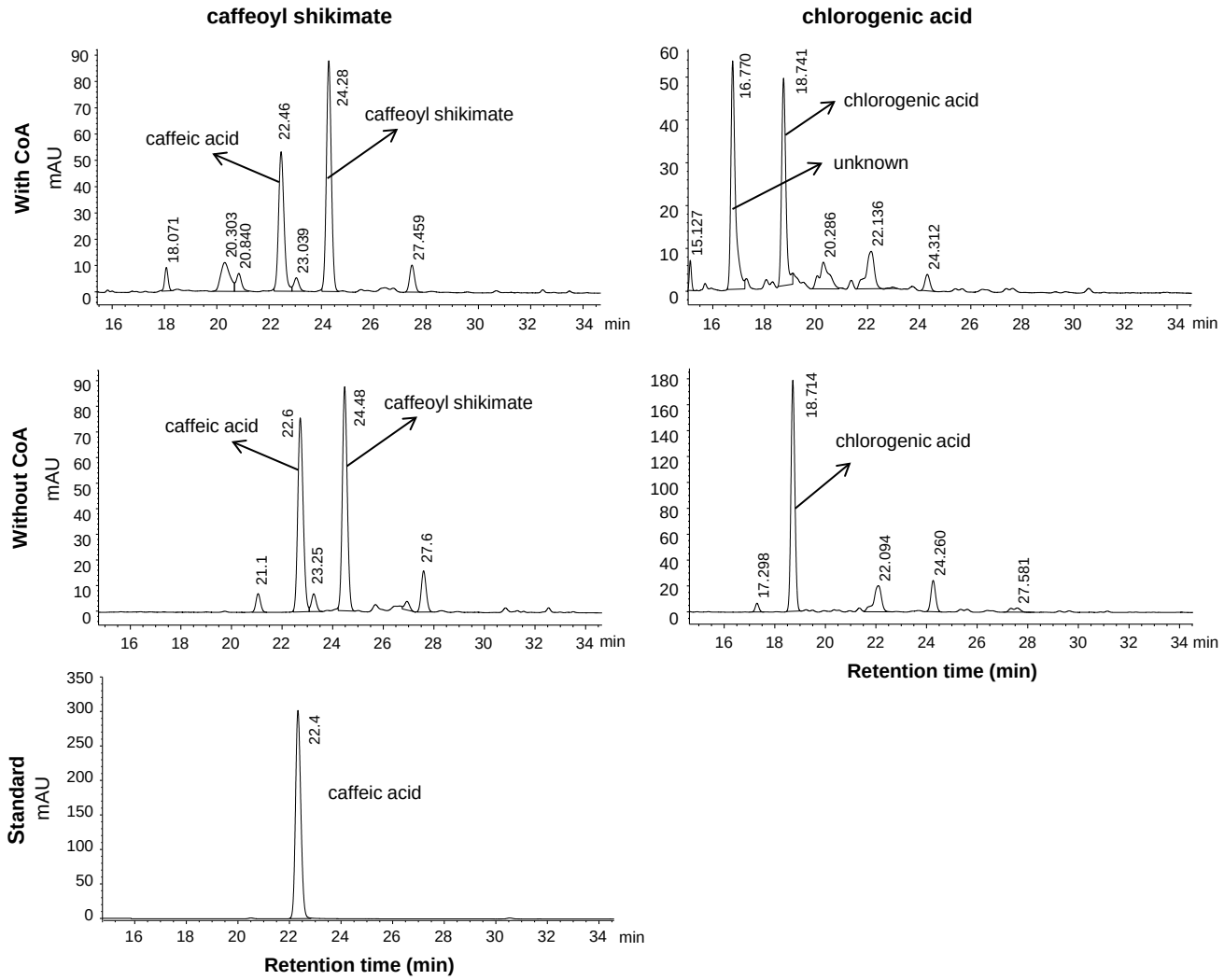

Fig. 4 Chromatograms of final products after incubating caffeoyl shikimate or chlorogenic acid with crude protein extracts of switchgrass in the presence or absence of CoA

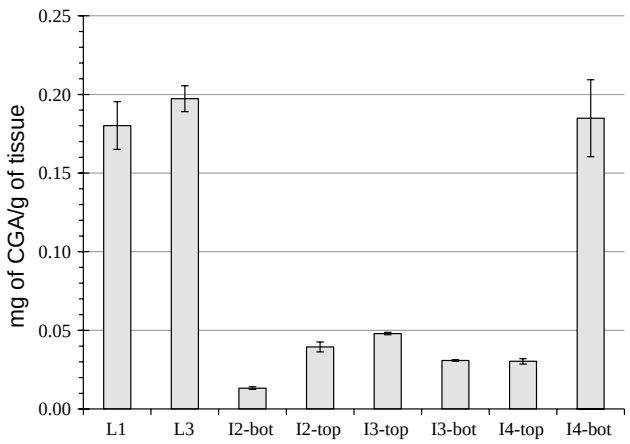

Fig. 5 CGA content (mg/gram fresh weight) in different tissues of E4 stage switchgrass St2 plants: L1 and L3, leaves 1 and 3; I2-top, I2-bot, I3-top, I3-bot, I4-top, I4-bot, *top* and *bottom* parts of internodes 2, 3 and 4

a developing internode, for example I4 of the E4 stage plant. Thus, the transcript levels of *PvHCT-Like1* in I4-bot are tenfold higher than in I4-mid, in contrast to the transcript levels of *PvHCT1a* and *PvHCT2a* which are 3.6 and 4.5 fold higher respectively in I4-mid than I4-bot (Fig. 6). These data suggest that PvHCT-Like1 is more likely to be involved in CGA biosynthesis than PvHCT1a or 2a.

Discussion

Switchgrass accumulates high levels of CGA but does not appear to contain a classical *HQT* gene

Most previous studies on CGA biosynthesis, as well as the lignin pathway enzymes HCT and C3′H, have been performed in dicotyledonous plants. HQT is the key enzyme

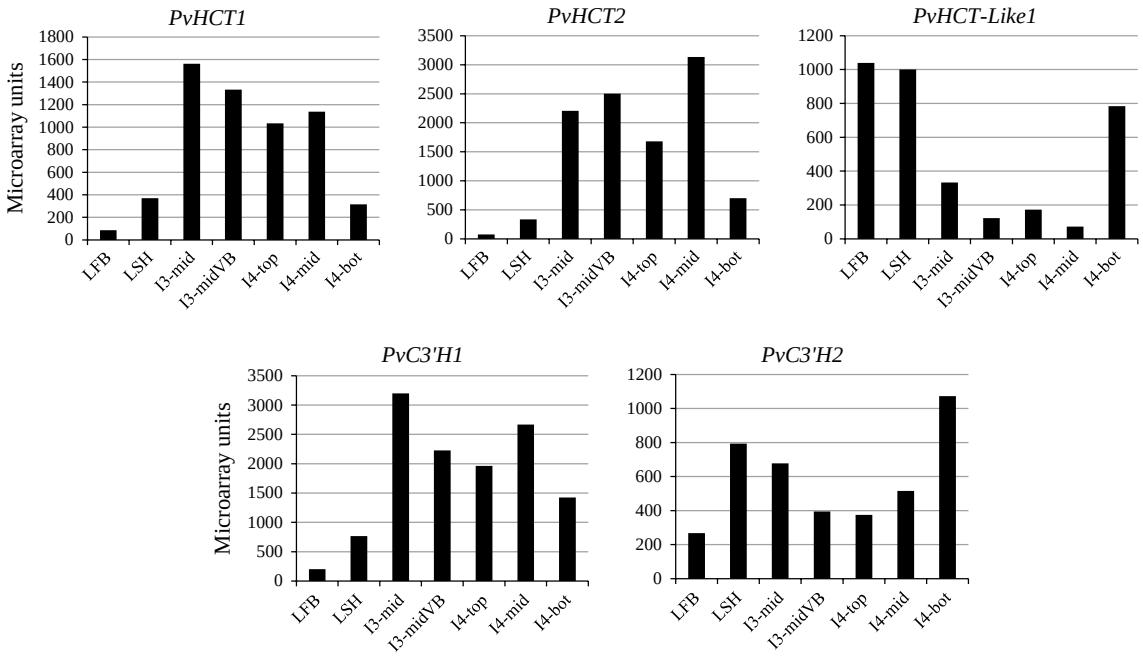

Fig. 6 Transcript data for *PvHCT1*, *PvHCT2*, *PvHCT-L1a*, *PvC3H1* and *PvC3′H2* obtained from the switchgrass gene expression atlas (PviGEA) web server (http://switchgrassgenomics.noble.org/). Data extracted and plotted were from E4 stage plants: Leaf blade (LFB); leaf sheath (LSH); middle part of internode 3 (I3-mid); the vascular bundles of I3-mid (I3-midVB); *top*, *middle* and *bottom* parts of internode 4 (I4-top, I4-mid, I4-bot)

involved in CGA biosynthesis in dicots (Niggeweg et al. 2004, Sonnante et al. 2010). However, although switchgrass accumulates high concentrations of CGA, the closest related switchgrass proteins to HQT in our phylogentic analysis were PvHCT1a and PvHCT2a. Furthermore, the closest related proteins to HQTs from other monocot species were also more closely related to HCTs than to HQTs. These results suggest that monocots that accumulate CGA do so without the involvement of a classical, dedicated HQT enzyme.

CGA levels and gene expression in elongating internodes

There is a gradation of lignin level in switchgrass young internodes, from the top (highest) to the bottom part, while CGA content shows the opposite trend, higher in the bottom than in the top (Shen et al. 2009). In this work we confirm this trend in internode 4 (the youngest) of E4 stage plants, where the base part has higher CGA content than the top part. Internodes 2 and 3 of E4 stage plants are much more lignified and have considerably lower CGA content.

According to the data extracted from PviGEA, the transcript levels of *PvHCT1a*, *PvHCT2a*, and *PvC3′H1* are higher in tissue that is more lignified (I3-mid, I3-midVB, I4-top, I4-mid) whereas the transcript levels of *PvHCT-Like1* and *PvC3′H2* are higher in less lignified tissue (LFB, LSH and I4-bot, with the exception of *PvC3′H2* in LFB). Concentrating our attention to a single developing internode (I4 of E4 stage plant), the bottom part (I4-bot) which is less lignified and has about sixfold higher CGA level than I4-top is also the tissue where *PvHCT-Like1* and *PvC3′H2* have the highest transcript levels and where *PvHCT1a*, *PvHCT2a* and *PvC3′H1* have the lowest transcript levels within the internode. This suggests that *PvHCT1a*, *PvHCT2a* and *PvC3′H1* could be related to lignin biosynthesis whereas *PvHCT-Like1* and *PvC3′H2* could be related to CGA biosynthesis. Unequivocal support for this hypothesis will require stable transformation of switchgrass with RNAi or antisense constructs for down-regulation of these genes.

Potential pathways to CGA and lignin in switchgrass

Figure 7 summarizes possible routes to lignin and CGA involving the reactions and enzymes investigated in the present work. Our results suggest that, for formation of caffeoyl CoA, 4-coumaroyl CoA is converted to 4-coumaroyl shikimate by HCT, this shikimate ester is hydroxylated by C3′H, and the caffeoyl shikimate generated could theoretically be further metabolized by two routes; conversion to caffeoyl CoA by an HCT enzyme acting in the reverse direction, or hydrolysis by a CSE enzyme to generate caffeic acid followed by conversion to caffeoyl CoA by 4CL, bypassing the reverse HCT reaction (Vanholme

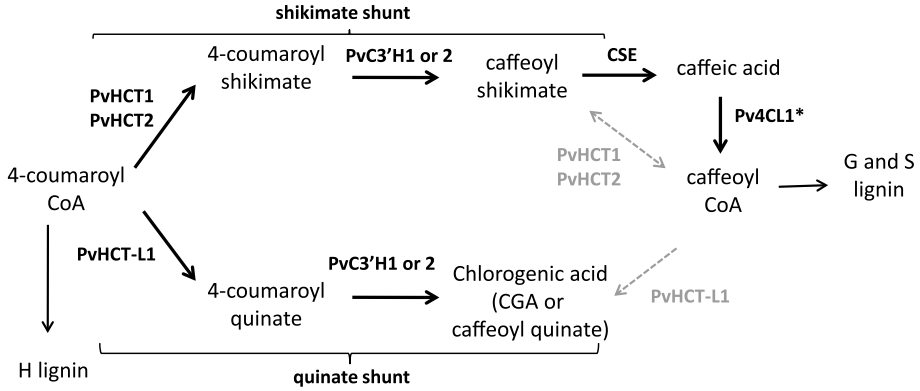

Fig. 7 Proposed scheme of routes for CGA and caffeoyl CoA biosynthesis in switchgrass. *Black arrows* mean most likely routes based on properties of recombinant enzymes, activities present in crude plant extracts, and gene expression data, and *dotted gray arrows* mean less likely routes based on the same criteria. *Asterisk* data from Xu et al. 2011

et al. 2013). Enzyme kinetic parameters of PvHCT1a and PvHCT2a indicate their preference towards shikimic acid as acyl acceptor. The most efficient step for both enzymes is the forward reaction converting 4-coumaroyl CoA to 4-coumaroyl shikimate (Kcat/Km 3.1 and 1.78 s^{-1} μM^{-1} respectively); they are also able to catalyze the reverse reaction converting caffeoyl shikimate to caffeoyl CoA although inefficiently (Kcat/Km = 0.123 or 0.09 $s^{-1}\mu M^{-1}$ respectively). This would complete the classical two-step involvement of HCT in the route called the shikimate shunt (Hoffmann et al. 2003). The inefficiency of the recombinant enzymes in carrying out this reverse reaction, and the fact that plant extracts are not able to carry out the generation of caffeoyl CoA from caffeoyl shikimate in the presence of CoA but instead hydrolyze the shikimate ester, suggests that the classical two-step involvement of HCT in the shikimate shunt is unlikely to occur in switchgrass (Hoffmann et al. 2003), and that a more likely pathway will involve a switchgrass ortholog of the recently discovered caffeoyl shikimate esterase (CSE) involved in lignin biosynthesis in Arabidopsis (Vanholme et al. 2013). Recombinant Pv4CL1 is readily able to generate caffeoyl CoA (Xu et al. 2011), thus providing the most likely route to this critical intermediate of both lignin and CGA formation in switchgrass. Additional evidence supporting this route is that neither recombinant enzymes nor plant extracts catalyze the formation of caffeoyl CoA from CGA.

Caffeoyl CoA is subsequently utilized for formation of lignin (G or S) or CGA. From caffeoyl CoA, CGA biosynthesis can be completed with one more step which could be catalyzed by PvHCT-Like1. Recombinant PvHCT-Like1 (different from PvHCT1a and PvHCT2a) showed preference towards quinic acid as acyl acceptor, so it is able to catalyze the formation of CGA from caffeoyl CoA, although less efficiently (Kcat/Km = 0.154 s^{-1} μM^{-1}) than the reaction which uses 4-coumaroyl CoA and quinic acid to generate 4-coumaroyl quinate (Kcat/Km = 0.74 s^{-1} μM^{-1}). The formation of 4-coumaroyl quinate was also observed using crude plant extracts. In this route, the biosynthesis of CGA would be completed by a C3′H enzyme which had the ability to hydroxylate 4-coumaroyl quinate. Recombinant PvC3′H1 and PvC3′H2 are able to hydroxylate both 4-coumaroyl shikimate and 4-coumaroyl quinate to generate caffeoyl shikimate or CGA, respectively, and could therefore be involved in both lignin and/or CGA biosynthesis.

Down-regulation of *HCT* or *C3′H* genes in dicot plants leads to reduction in lignin content, but also negatively impacts growth (Hoffmann et al. 2004; Reddy et al. 2005; Chen et al. 2006). Studies are in progress to address the involvement of the switchgrass HCT and C3′H orthologs in the formation of lignin and CGA through transgenic approaches. This is complicated, however, by the presence of the multiple *HCT* and *C3′H* genes in switchgrass, and the need to therefore generate both single knock-down lines, and lines in which two or more related genes are targeted. These studies should ultimately determine whether switchgrass uses separate redundant or shared parallel pathways for the biosynthesis of its major phenylpropanoid-derived end-products.

Materials and methods

Plant material and reagents

A single seedling from the Alamo cultivar of *Panicum virgatum* (switchgrass) was randomly selected, clonally propagated, and plants grown under standard greenhouse conditions.

The HCT substrates 4-coumaroyl CoA and caffeoyl CoA were synthesized as described previously (Stockigt and Zenk, 1975). Shikimic, chlorogenic and quinic acids were purchased from Sigma-Aldrich (St. Louis, MO). 4-Coumaroyl shikimate and caffeoyl shikimate were kindly provided by Dr. John Ralph (University of Wisconsin-Madison). The vector pYeDP60 and the *Saccharomyces cerevisiae* strain WAT11 were provided by Dr. Denis Pompon (Centre National de la Recherche Scientifique, France).

Cloning of cDNAs

Full length cDNA sequences found in switchgrass public databases (http://switchgrassgenomics.noble.org/ or http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=switchgrass) were used for primer design to clone targeted genes. Total RNA was isolated from leaf or stem using Trizol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's instructions, and cDNA synthesis was performed using the SuperScript III First-Strand System for RT-PCR Kit (Invitrogen) using an oligo-dT primer. Open reading frames were amplified by PCR using KOD DNA polymerase (EMD, San Diego, CA), and cloned into the *E. coli* expression vectors pET28a (EMD) or *Saccharomyces cerevisiae* expression vector pYeDP60 using regular cloning, or into pDEST17 using the Gateway technology (Invitrogen). The primers used for cloning are listed in Supplemental Table S3. Plasmids were isolated using the Wizard Plus Minipreps DNA Purification System (Promega, Madison, WI), and their insert cDNAs were sequenced in both directions for verification.

Phylogenetic analysis

Multiple sequence alignment (MSA) on the full length protein sequences was employed to perform phylogeny reconstruction. MAFFT (Katoh et al. 2005) was used in the alignment by employing the highly accurate method L-INS-I (Ahola et al. 2006; Nuin et al. 2006). Approximate maximum likelihood (ML) trees were built using FastTree 2 (Price et al. 2010) with default parameters.

Expression of switchgrass HCTs in *E. coli*

E. coli Rosetta strain cells harboring the target HCT constructs were cultured at 37 °C until OD_{600} reached 0.7-0.9, and isopropyl 1-thio β-galactopyranoside (IPTG) was added to a final concentration of 0.5 mM to induce the heterologous protein expression. The culture was incubated at 16 °C for 18-20 h, and tubes with 25 mL of culture were spun down to collect the pellets which were frozen at -80 °C. Pellets were thawed, resuspended in 2 mL of extraction-washing buffer (10 mM imidazole, 50 mM Tris–HCl pH 8.0, 500 mM NaCl, 10 % glycerol and 10 mM β-mercaptoethanol) and sonicated three times for 20 s. The supernatants were recovered after centrifugation (16,000×*g*), and equilibrated. Ni–NTA beads (Qiagen, Germantown, MD) were added to allow the His-tagged proteins to bind to the beads. The suspension was incubated at 4 °C for 30 min under constant inversion and unbound proteins were washed away three times with 1 mL of extraction-washing buffer. Target proteins were eluted with 250 μL of elution solution (250 mM imidazole, 50 mM Tris–HCl buffer pH 8.0, 500 mM NaCl, 10 % glycerol and 10 mM β-mercaptoethanol). Their purity was verified by SDS-PAGE, and protein concentrations were determined using the BioRad protein assay (BioRad, Hercules, CA).

Expression of switchgrass C3′H in *Saccharomyces cerevisiae*

The *PvC3′H1* and *PvC3′H2* cDNAs were cloned between the *Bam*HI and *Eco*RI restriction enzyme sites of *Saccharomyces cerevisiae* expression vector pYeDP60 (Pompon et al. 1996) The *S. cerevisiae* strain WAT11 was transformed, cultured and induced as described previously (Pompon et al. 1996). After 30 h of induction, cell pellets were recovered and yeast microsomes were prepared as described previously (Liu et al. 2003). Microsomes were used for C3′H enzyme activity assays as described below.

Preparation of switchgrass protein extracts

From 0.3 to 0.5 g of switchgrass tissue was ground very finely using a freezing mill, and a crude protein extract prepared as described previously (Gallego-Giraldo et al. 2011).

Assay of HCT enzyme activity and determination of kinetics

Five to 100 ng of pure recombinant PvHCT1a, PvHCT2a and PvHCT-L1 proteins, or 5–10 μg of plant protein extracts, were incubated at 30 °C for 10–60 min with 100 mM sodium phosphate buffer pH 7.5, 500 μM shikimic acid or quinic acid, 500 μM dithiothreitol (Roche, Madison, WI) and 2–80 μM acyl donor substrate (4-coumaroyl CoA or caffeoyl CoA) in a final volume of 100 μL. The reverse reactions were tested using caffeoyl shikimate or caffeoyl quinate (2–80 μM) and 500 μM CoA. The reactions were terminated by adding 10 μL of glacial acetic acid. Reaction products were injected onto an HPLC with a reverse-phase C18 column (Spherisorb 5μ ODS2, Waters, Milford, MA) and separated in a step gradient using 1 % phosphoric acid in water as solvent A and acetonitrile as solvent B. To quantify products, peak areas were measured and converted to units of quantity using calibration curves that were constructed with authentic standards of each product except for 4-coumaroyl quinate, which was not available; in this latter case, 4-coumaroyl shikimate was used as standard .

Determination of C3′H enzyme activity

C3′H activity assays used 2–10 μL of recombinant yeast microsomes, 100 mM sodium phosphate buffer pH 7.5,

600 μM NADPH and 10 or 40 μM 4-coumaroyl shikimate. Since the substrate 4-coumaroyl quinate was not commercially available, it was prepared in situ using 100 ng of PvHCT1a enzyme, 100 mM sodium phosphate buffer pH 7.5, 500 μM quinic acid, 500 μM dithiothreitol and 200 μM 4-coumaroyl CoA, with incubation for 1 or 2 h. For this coupled reaction, NADPH was added to 600 μM final concentration. For assays with both 4-coumaroyl shikimate and 4-coumaroyl quinate, 2–10 μL of microsomes were then added (along with additional buffer to maintain 100 mM in the coupled assays), and the reactions were incubated for an additional 30–60 min at 30 °C. After addition of 10 μL of glacial acetic acid, the mixtures were spun down and the supernatants analyzed by reverse-phase HPLC on a C18 column as described above.

Quantification of CGA

Finely ground switchgrass tissue (20–30 mg, prepared using a freezing mill) were exactly weighed and 1 mL of 50 % v/v methanol containing 1.5 % v/v acetic acid was added and mixed by vortexing. The suspension was incubated at room temperature overnight with constant rotation. After centrifugation, 80 μL of the supernatant was analyzed by reverse-phase HPLC as described above.

Analysis of transcript data from the switchgrass gene expression atlas

We utilized the publicly available switchgrass gene expression atlas PviGEA (Zhang et al. 2013) (http://switchgrassgenomics.noble.org/) to investigate the transcript data for our five genes as obtained by Affymetrix microarray analysis. By blast queries of our cDNAs sequences, we identified the probe set IDs: *PvHCT1a*: AP13ISTG44531, *PvHCT2a*: AP13CTG44233-Rc, *PvHCT-Like1*: AP13ISTG74189, *PvC3′H1a*: AP13ISTG41630 and *PvC3′H2*: AP13ISTG55845, which represent the switchgrass unitranscrips (PviUT). Using the probe set IDs, we extracted and plotted the transcript data in the desired tissue.

Acknowledgments We thank Drs. Jerome Verdier and Lina Gallego-Giraldo for critical reading of the manuscript. This work was supported by the BioEnergy Science Center, a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

References

Ahola V, Aittokallio T, Vihinen M, Uusipaikka E (2006) A statistical score for assessing the quality of multiple sequence alignments. BMC Bioinformat 7:484

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol 25:759–761

Chen F, Srinivasa Reddy MS, Temple S, Jackson L, Shadle G, Dixon RA (2006) Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (*Medicago sativa* L.). Plant J 48:113–124

Comino C, Lanteri S, Portis E, Acquadro A, Romani A, Hehn A, Larbat R, Bourgaud F (2007) Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. BMC Plant Biol 7:14

Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E (2009) The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. BMC Plant Biol 9:30

Davison BH, Drescher SR, Tuskan GA, Davis MF, Nghiem NP (2006) Variation of S/G ratio and lignin content in a Populus family influences the release of xylose by dilute acid hydrolysis. Appl Biochem Biotechnol 129–132:427–435

Escamilla-Trevino LL, Shen H, Uppalapati SR, Ray T, Tang Y, Hernandez T, Yin Y, Xu Y, Dixon RA (2010) Switchgrass (*Panicum virgatum*) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties. New Phytol 185:143–155

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M Jr, Chen F, Foston M, Ragauskas A, Bouton J, Dixon RA, Wang Z-Y (2011) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci USA 108:3803–3808

Gallego-Giraldo L, Escamilla-Trevino L, Jackson LA, Dixon RA (2011) Salicylic acid mediates the reduced growth of lignin down-regulated plants. Proc Natl Acad Sci USA 108:20814–20819

Grabber JH (2005) How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. Crop Sci 45:820–831

Hoffmann L, Maury S, Martz F, Geoffroy P, 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

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

Jackson LA, Shadle GL, Zhou R, Nakashima J, Chen F, Dixon RA (2008) Improving saccharification efficiency of alfalfa stems through modification of the terminal stages of monolignol biosynthesis. Bioenergy Res 1:180–192

Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–518

Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J (2007) Chlorogenic acid synthesis in coffee: an analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. Plant Sci 172:978–996

Liu CJ, Huhman D, Sumner LW, Dixon RA (2003) Regiospecific hydroxylation of isoflavones by cytochrome P450 81E enzymes from *Medicago truncatula*. Plant J 36:471–484

Mahesh V, Million-Rousseau R, Ullmann P, Chabrillange N, Bustamante J, Mondolot L, Morant M, Noirot M, Hamon S, de Kochko A (2007) Functional characterization of two *p*-coumaroyl ester 3′-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. Plant Mol Biol 64:145–159

McLaughlin SB, Adams Kszos L (2005) Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass Bioenergy 28:515–535

Moglia A, Comino C, Portis E, Acquadro A, De Vos RCH, Beekwilder J, Lanteri S (2009) Isolation and mapping of a C3′H gene (CYP98A49) from globe artichoke, and its expression upon UV–C stress. Plant Cell Rep 28:963–974

Moore KJ, Moser LE, Vogel KP, Waller SS, Johnson BE, Pedersen JF (1991) Describing and quantifying growth stages of perennial forage grasses. Agron J 83:1073–1077

Morant M, Schoch GA, Ullmann P, Ertunç T, Little D, Olsen CE, Petersen M, Negrel J, Werck-Reichhart D (2007) Catalytic activity, duplication and evolution of the CYP98 cytochrome P450 family in wheat. Plant Mol Biol 63:1–19

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat Biotechnol 22:746–754

Nuin PA, Wang Z, Tillier ER (2006) The accuracy of several multiple sequence alignment programs for proteins. BMC Bioinformat 7:471

Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450 s in optimized redox environments. Meth Enzymol 272:51–64

Price MN, Dehal PS, Arkin AP (2010) FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS ONE 5:e9490

Ralph J, Akiyama T, Kim H, Lu F, Schatz PF, Marita JM, Ralph SA, Reddy MS, Chen F, Dixon RA (2006) Effects of coumarate 3′-hydroxylase down-regulation on lignin structure. J Biol Chem 281:8843–8853

Reddy MS, Chen F, Shadle G, Jackson L, Aljoe H, Dixon RA (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). Proc Natl Acad Sci USA 102:16573–16578

Schmer MR, Vogel KP, Mitchell RB, Perrin RK (2008) Net energy of cellulosic ethanol from switchgrass. Proc Natl Acad Sci USA 105:464–469

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from *Arabidopsis thaliana* is a 3′-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J Biol Chem 276:36566–36574

Shen H, Fu CX, Xiao XR, Ray T, Tang YH, Wang ZY, Chen F (2009) Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. Bioenergy Res 2:233–245

Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DGJ, Wang H, Jackson L, Tang Y, Neal Stewart C Jr (2011) Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. New Phytol 193:121–136

Shen H, Mazarei M, Rudis MR, Tang YH, Hisano H, Jackson L, Li G, Hernandez T, Chen F, C. Stewart Jr N, Wang Z, Dixon RA (2013). A genomic approach to deciphering pathways for lignin biosynthesis in switchgrass (*Panicum virgatum* L.). Plant Cell, in revision

Sonnante G, D'Amore R, Blanco E, Pierri CL, De Palma M, Luo J, Tucci M, Martin C (2010) Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiol 153:1224–1238

Stockigt J, Zenk MH (1975) Chemical syntheses and properties of hydroxycinnamoyl-Coenzyme A derivatives. Z Naturforsch [C] 30:352–358

Talukder K (2006) Low-lignin wood–a case study. Nat Biotechnol 24:395–396

van der Rest B, Danoun S, Boudet AM, Rochange SF (2006) Down-regulation of cinnamoyl-CoA reductase in tomato (*Solanum lycopersicum* L.) induces dramatic changes in soluble phenolic pools. J Exp Bot 57:1399–1411

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

Villegas RJ, Kojima M (1986) Purification and characterization of hydroxycinnamoyl D-glucose: quinate hydroxycinnamoyl transferase in the root of sweet potato, *Ipomoea batatas* Lam. J Biol Chem 261:8729–8733

Wullschleger SD, Davis EB, Borsuk ME, Gunderson CA, Lynd LR (2010) Biomass production in switchgrass across the United States: database description and determinants of yield. Agron J 102:1158–1168

Xu B, Escamilla-Trevino LL, Noppadon S, Shen Z, Shen H, Percival Zhang YH, Dixon RA, Zhao B (2011) Silencing of 4-coumarate: coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. New Phytol 92:611–625

Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN (2008) Plants to power: bioenergy to fuel the future. Trends Plant Sci 13:421–429

Zhang J-Y, Lee Y-C, Torres-Jerez I, Wang M, Yin Y, Chou W-C, He J, Shen H, Srivastava AC, Pennacchio C, Lindquist E, Grimwood J, Schmutz J, Xu Y, Sharma M, Sharma R, Bartley LE, Ronald PC, Saha MC, Dixon RA, Tang Y, Udvardi MK (2013) Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). Plant J 74:160–173