2012

Brown midrib2 (Bmr2) encodes the major 4-coumarate: coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench)

Ana Saballos
University of Florida

Scott Sattler
USDA-ARS, Scott.Sattler@ars.usda.gov

Emiliano Sanchez
Washington State University

Timothy P. Foster
University of Florida

Zhanguo Xin
USDA-ARS, zhanguo.xin@ars.usda.gov

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2012) **70**, 818–830 doi: 10.1111/j.1365-313X.2012.04933.x

Brown midrib2 (*Bmr2*) encodes the major 4-coumarate: coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench)

Ana Saballos[1], Scott E. Sattler[2], Emiliano Sanchez[3], Timothy P. Foster[1], Zhanguo Xin[4], ChulHee Kang[3], Jeffrey F. Pedersen[2] and Wilfred Vermerris[1,*]

[1]*Agronomy Department and Genetics Institute, University of Florida, Gainesville, FL 32610, USA,*
[2]*US Department of Agriculture – Agricultural Research Service, Grain Forage and Bioenergy Research Unit, Lincoln, NE 68583, USA, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA,*
[3]*School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA, and*
[4]*US Department of Agriculture – Agricultural Research Service, Lubbock, TX 79415, USA*

Received 7 October 2011; revised 11 January 2012; accepted 27 January 2012; published online 4 April 2012.
*For correspondence (e-mail wev@ufl.edu).

SUMMARY

Successful modification of plant cell-wall composition without compromising plant integrity is dependent on being able to modify the expression of specific genes, but this can be very challenging when the target genes are members of multigene families. 4-coumarate:CoA ligase (4CL) catalyzes the formation of 4-coumaroyl CoA, a precursor of both flavonoids and monolignols, and is an attractive target for transgenic down-regulation aimed at improving agro-industrial properties. Inconsistent phenotypes of transgenic plants have been attributed to variable levels of down-regulation of multiple *4CL* genes. Phylogenetic analysis of the sorghum genome revealed 24 4CL(-like) proteins, five of which cluster with bona fide 4CLs from other species. Using a map-based cloning approach and analysis of two independent mutant alleles, the sorghum *brown midrib2* (*bmr2*) locus was shown to encode 4CL. *In vitro* enzyme assays indicated that its preferred substrate is 4-coumarate. Missense mutations in the two *bmr2* alleles result in loss of 4CL activity, probably as a result of improper folding as indicated by molecular modeling. *Bmr2* is the most highly expressed *4CL* in sorghum stems, leaves and roots, both at the seedling stage and in pre-flowering plants, but the products of several paralogs also display 4CL activity and compensate for some of the lost activity. The contribution of the paralogs varies between developmental stages and tissues. Gene expression assays indicated that *Bmr2* is under auto-regulatory control, as reduced 4CL activity results in over-expression of the defective gene. Several *4CL* paralogs are also up-regulated in response to the mutation.

Keywords: 4-coumarate:CoA ligase, *bmr*, *brown midrib*, cell wall, lignin, sorghum.

INTRODUCTION

Cell-wall biogenesis is a complex process involving a large number of enzymes of various classes that together synthesize monomers for the various cell-wall polymers, coordinate polymer assembly and deposition, and modify the wall in a developmentally regulated fashion (Yokoyama and Nishitani, 2004; Penning *et al.*, 2009). Detailed understanding of the specific roles of these various enzymes, including their kinetic parameters and spatio-temporal control, is critical in order to be able to tailor the composition of the cell wall to specific downstream processes, including use of cell walls for food, fodder, fibers and fuel.

The cell-wall polymer lignin has received considerable attention because of its negative effect on the conversion efficiency of lignocellulosic biomass to fermentable sugars that can be used for production of fuels and chemicals (Chen and Dixon, 2007; Vermerris *et al.*, 2007). Lignin is a phenolic polymer that is formed via oxidative coupling of hydroxycinnamyl alcohols (Ralph *et al.*, 2004) (Figure S1). Genetic and transgenic approaches have been used in a number of species to reduce lignin content and modify lignin subunit composition (Vanholme *et al.*, 2008). In some cases, perturbations of lignin biosynthesis have been shown to have

serious repercussions on plant fitness (Jones *et al.*, 2001; Nakashima *et al.*, 2008; Vermerris *et al.*, 2010), but in other cases no apparent detrimental effects were observed (Chabannes *et al.*, 2001; Jackson *et al.*, 2008). Hence, detailed and comprehensive knowledge of lignin biosynthetic enzymes is necessary to improve the quality of the feedstock without compromising plant fitness.

4-coumarate:coenzyme A ligase (4CL; EC 6.2.1.12) is an obvious target for down-regulation, because reduced 4CL activity is expected to decrease the flux of monolignol precursors and consequently lignin content. Using 4-coumarate as a substrate, 4CL catalyzes the formation of 4-coumaroyl CoA via a two-step ATP- and Mg^{2+}-dependent mechanism that first leads to a hydroxycinnamate-AMP anhydride intermediate, followed by nucleophilic substitution of AMP by S-CoenzymeA (Knobloch and Hahlbrock, 1975; Hu *et al.*, 2010) (Figure S1). Other members of the 4CL family are also able to convert cinnamate, caffeate, ferulate, 5-hydroxyferulate and sinapate to their corresponding CoA thiol esters *in vitro* (Knobloch and Hahlbrock, 1975; Lee and Douglas, 1996; Allina *et al.*, 1998; Cukovic *et al.*, 2001; Ehlting *et al.*, 2001; Hamberger and Hahlbrock, 2004). Several 4CL isoforms with different affinities towards the possible substrates have been identified, leading to the hypothesis that variation in expression of the genes encoding the various 4CL isoforms plays a role in the regulation of lignin content and composition throughout development of the plant (Ehlting *et al.*, 1999; Lindermayr *et al.*, 2002; Hamberger and Hahlbrock, 2004). In addition, some of these isoforms are involved in the biosynthesis of 4-coumaroyl CoA-derived flavonoids (Hamberger and Hahlbrock, 2004; Xu *et al.*, 2009). Phylogenetic analysis of 4CL sequences does indeed support the existence of two distinct classes of enzyme (Hu *et al.*, 1998; Ehlting *et al.*, 1999; Cukovic *et al.*, 2001).

To date, down-regulation of *4CL* expression has relied on transgenic approaches. Transgenic tobacco (*Nicotiana tabacum*; Kajita *et al.*, 1997), Arabidopsis (Lee *et al.*, 1997), *Populus* (Hu *et al.*, 1999; Voelker *et al.*, 2010), switchgrass (*Panicum virgatum*; Xu *et al.*, 2011) and rice (*Oryza sativa*; Gui *et al.*, 2011) in which *4CL* had been down-regulated all displayed reductions in lignin content, but the effect on lignin subunit composition varied considerably. It is possible that the variation in phenotype of these transgenic plants was due to variable levels of down-regulation among the various members of the *4CL* family.

Phylogenetic analyses revealed the existence of multiple *4CL* genes in maize (*Zea mays*) and rice whose expression patterns vary depending on the tissue and developmental stage (Guillaumie *et al.*, 2007; Penning *et al.*, 2009). Perturbations in the monolignol biosynthetic pathway in grasses can lead to brown vascular tissue in the leaves and stems, as apparent in the *brown midrib* mutants of maize, sorghum (*Sorghum bicolor*) and pearl millet (*Pennisetum glaucum*; Sattler *et al.*, 2010). To date, four allelic classes of sorghum *bmr* mutants have been identified, represented by the mutants *bmr2*, *bmr6*, *bmr12* and *bmr19*, with the first three showing the greatest potential for increasing biomass conversion (Saballos *et al.*, 2008). The *Bmr12* and *Bmr6* genes have been cloned (Bout and Vermerris, 2003; Saballos *et al.*, 2009; Sattler *et al.*, 2009). The lignin of the *bmr2* mutant shows reductions in both guaiacyl and syringyl residues (Saballos *et al.*, 2008). We show here that the *bmr2* mutation is located in one of the sorghum *4CL* genes, and that the *bmr2* mutant phenotype provides evidence for a major but not exclusive role for this *4CL* gene in lignification.

RESULTS

Chemical characterization of an independent *bmr2* mutant

The original *bmr2* mutant described by Porter *et al.* (1978) was designated as *bmr2-ref* by Saballos *et al.* (2008). After a complementation test, an independent *bmr2* allele was discovered in the TILLING population described by Xin *et al.* (2008), and is designated here as the *bmr2-2* allele. Both *bmr2* mutants displayed the characteristic brown coloration of the midrib (Figure S2). In addition, sclerenchyma tissue in the stems and leaves was brown, and displayed minimal staining with Wiesner reagent and a slightly reduced level of autofluorescence, indicative of changes in lignin composition (Figure S3). No apparent differences in staining with the Wiesner reagent or autofluorescence were observed in the protoxylem of developing *bmr2* stems, in contrast to leaf xylem (Figure S3). Comparisons of the composition of midribs and ground stover from wild-type BTx623 and *bmr2-2* plants were consistent with the changes reported by Saballos *et al.* (2008) for the *bmr2-ref* mutant, with modest reductions in the intensity of most guaiacyl (G) residues, and a limited reduction in syringyl (S) residues (Figures S4–S6). The Klason lignin concentrations of *bmr2-2* stover and leaf tissue were 17 and 9 mg g^{-1}, respectively, representing a 20% reduction compared to the wild-type in both cases. GC-MS analysis of methanol-soluble metabolites revealed accumulation of free ferulic acid in stems but not leaves of *bmr2* plants (Figure S7). In contrast, no differences were found in the concentration of free 4-coumaric acid in stems or leaves. Small variations in the concentrations of approximately 200 other metabolites did not show consistent correlation with the presence of the mutation.

Mapping and identification of the *Bmr2* gene

To identify the gene responsible for the *bmr2* phenotype, a map-based cloning approach was undertaken using an F_2 population generated from crossing line AMP11, which carries the *bmr2-ref* allele, and the unrelated cultivar 'Theis'. The map location of *bmr2* was determined to be 2.14 cM away from SSR marker Xtxp26. The SSR marker Xtxp26, previously undetermined, was located on chromosome 4

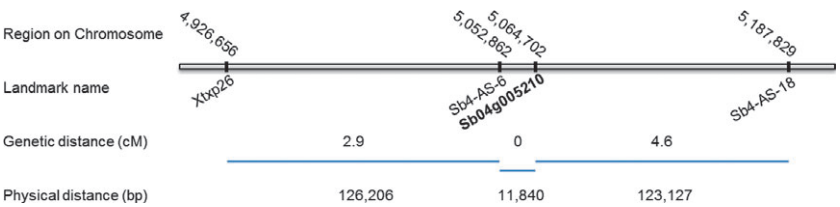

Figure 1. Genomic location of *Bmr2* and linked molecular markers on sorghum chromosome 4. *Bmr2* (Sb04g005210) is shown in bold. The base pair numbers correspond to the *Sorghum bicolor* genomic database version 7.0. (www.phytozome.net/sorghum)

based on the sorghum genome sequence. Seventy homozygous F_2 plants and 20 homozygous *bmr* F_3 plants were genotyped using newly designed molecular makers located within a 4 Mb region around Xtxp26 (Table S1), narrowing the interval containing the *bmr2* locus to 262 kb. No recombination events were observed between SSR marker SB4-AS-6 and the *bmr2* phenotype among 90 homozygous individuals (Figure 1).

Twenty-two annotated genes were present within this 262 kb interval, including Sb04g005210, a homolog of the Arabidopsis genes *4CL1*, *4CL2* and *4CL3*, located only 11.8 kb away from marker SB4-AS-6. Considering the changes in lignin composition in the *bmr2* mutant, Sb04g005210 was the most obvious candidate gene.

The candidate gene was PCR-amplified from the *bmr2* mutants and wild-type controls and sequenced. The *bmr2-ref* and *bmr2-2* alleles contained G → A transitions within the coding sequence, at positions +468 and +1419, respectively (based on the Sb04g005210 genomic sequence). These missense mutations led to the amino acid substitutions Gly111Asp and Gly262Arg, respectively. Both mutations created novel restriction sites compared with the reference genome sequence, which were detectable as CAPS markers (Figure 2). These results were confirmed using independently isolated genomic DNA from mutants and wild-type controls. In the absence of an amenable transformation system for sorghum, the demonstration that two independent alleles associated with the same mutant phenotype both contain mutations is considered proof that the *Bmr2* gene encodes 4CL.

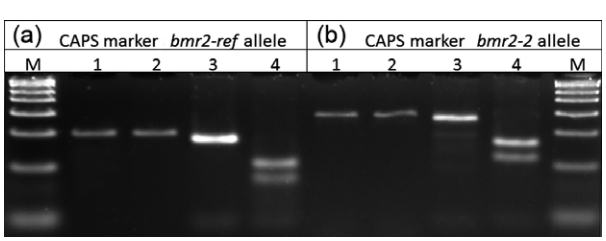

Figure 2. CAPS markers that distinguish between the wild-type and *bmr2* alleles.
(a) Lane 1, undigested PCR fragment for OK11 (wild-type). Lane 2, undigested PCR fragment for *bmr2-ref*. Lane 3, *Taq*I-digested wild-type PCR fragment. Lane 4, *Taq*I-digested *bmr2-ref* PCR fragment.
(b) Lane 1, undigested PCR fragment for BTx623 (wild-type). Lane 2, undigested PCR fragment for *bmr2-2*. Lane 3, *Mnl*I-digested wild-type PCR fragment. Lane 4, *Mnl*I-digested *bmr2-2* PCR fragment. M, DNA size markers.

Phylogenetic analysis of sorghum 4CL family members

The availability of the sorghum genome sequence allows identification of all *4CL* gene family members, which is the initial step in understanding their roles in sorghum. 4CL is a member of the acyl:CoA synthetase family, which is functionally divided into several sub-families (Khurana *et al.*, 2010). A search of the sorghum genome revealed 24 sequences with similarity to Arabidopsis *4CL* genes (Costa *et al.* 2005; Table S2A). At the amino acid level, the sequences showed identities ranging from 19.3 to 59.5% with At4CL1 and 19.4–61.4% with At4CL3. A phylogenetic analysis partitioned these 24 sequences into various clades, consistent with the dendrogram reported by Souza *et al.* (2008), in which all the proteins with reported 4CL activity grouped together in one clade, referred to as the bona fide 4CL clade. The members of this clade have been previously divided into classes I and II, which are involved in lignification and flavonoid production, respectively (Hu *et al.*, 1998). Four of the five sorghum sequences within the bona fide 4CL clade, including Bmr2, are classified as class I (Figure 3). Distinction between classes I and II was possible in both grasses and dicots, but not in any of the mosses, lycophytes and gymmosperms for which sequence data are available (Figure 3). This phylogenetic analysis indicated that the *4CL* genes in sorghum are organized in a manner that is similar overall to what has been observed in other angiosperms.

Activity of Bmr2 and paralogous 4CL enzymes

Bmr2 and the paralogous proteins encoded by Sb10g0026130, Sb07g007810 and Sb04g031010 were expressed in *Escherichia coli* as 6× His-tagged proteins, and purified from cell extracts using affinity chromatography on a nickel-containing resin. Their 4CL activity was determined using the substrates 4-coumarate, ferulate, caffeate, cinnamate and sinapate in a concentration range of 5–150 μM (Table 1 and Figure S8). Under our assay conditions, all enzymes showed the highest activity against the substrate 4-coumarate, and none displayed activity against sinapate. Bmr2 will preferentially convert 4-coumarate followed by ferulate, caffeate and cinnamate, consistent with activities reported for other bona fide 4CLs. Substrate inhibition was observed for caffeate at concentrations above 40 μM. The fivefold lower enzyme activity observed against cinnamate relative to 4-coumaric acid is consistent with the involvement of the 4-hydroxyl group in docking the phenylpropanoid

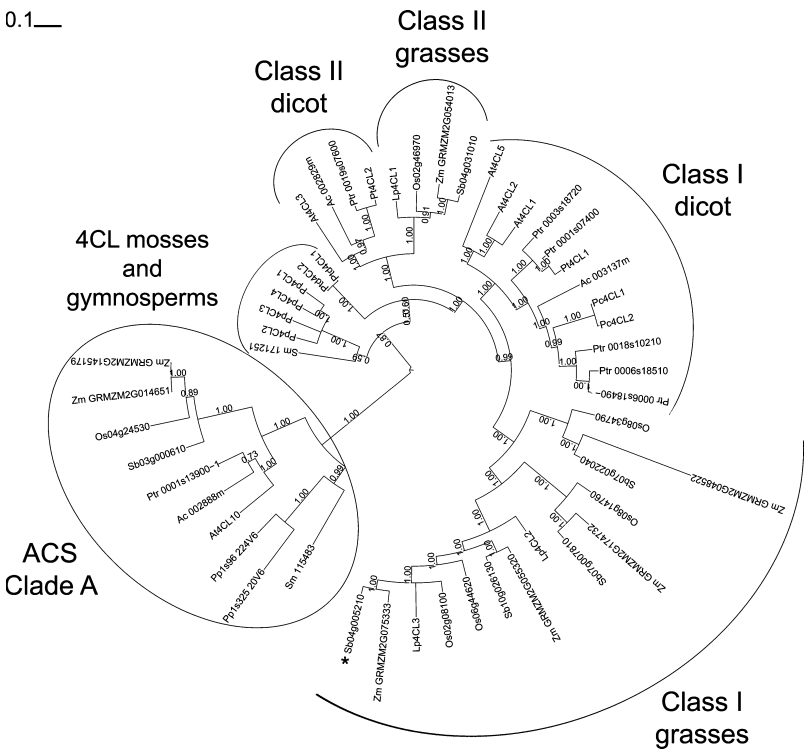

Figure 3. Phylogenic relationship of acyl CoA synthetases that cluster with proteins proven to have activity against phenylpropanoid substrates. The oval indicates clade A, identified by Souza *et al.* (2008), which is used as a root. The classes within the bona fide 4CL clade are indicated. Bmr2 is indicated by an asterisk. At4CL10 corresponds to AtACS5 described by Souza *et al.* (2008). Bootstrap values indicating the level of support for the displayed representation after re-sampling are shown on the branches. Branch lengths reflect the number of amino acid substitutions as indicated by the scale at the top left. GenBank accession numbers and Phytozome identifiers for the sequences used in this analysis are listed in Table S2B.

substrate within the substrate-binding pocket of the active site. The substrate preference and activity of Sb07g007810 were similar overall to that of Bmr2. The most obvious difference was that Sb07g007810 displayed a threefold higher activity against cinnamate than Bmr2, with no evidence for substrate inhibition within the concentration range tested. Even though 4-coumarate was also the preferred substrate of Sb10g0026130 and Sb04g031010, the activities of these two enzymes against 4-coumarate were three- and twofold lower, respectively, than that of Bmr2. Activity against caffeate and cinnamate was even lower, and subject to substrate inhibition at higher concentrations. (Table 1 and Figure S8). Activity against ferulate for both enzymes was very low and limited to narrow concentration ranges: 5–15 μM for Sb04g031010 and 25–40 μM for Sb10g0026130.

Immunoblots prepared from extracts of wild-type and *bmr2* stalks incubated with polyclonal antibody against recombinant Bmr2 showed only a weak signal in the *bmr2* extracts (Figure 4a). The difference between wild-type and mutants was not due to improper loading, based on the consistent signal strength observed for *S*-adenosylmethionine synthetase. The use of polyclonal antibodies and a distance of 141 amino acids between the two amino acid changes in the mutant proteins precluded the possibility that both mutations affect the same epitope of the protein.

To determine the effects that the *bmr2* mutations have on *in planta* 4CL activity, soluble proteins extracted from 6-week-old stalks were assayed for 4CL activity. Extracts from *bmr2* stalks had drastically reduced 4CL activity using 4-coumarate, caffeate and ferulate substrates relative to wild-type. No activity was detected in any extract using sinapate or cinnamate as substrates. The substrate preferences of plant extracts mirrored those observed with Bmr2 recombinant protein, but the activities of *bmr2-ref* and *bmr2-2* relative to the wild-type were 14 and 11% for 4-coumarate, 7 and 6% for ferulate, and 8 and 12% for caffeate, respectively (Figure 5). Differences in enzyme activity between the two mutants were not statistically significant, indicating that both mutations had similar effects on enzyme activity. To determine whether the mutant versions of Bmr2 had any residual activity, wild-type and Bmr2 proteins were expressed as thioredoxin fusion proteins in *E. coli*. Without addition of thioredoxin, Bmr2 mutant proteins were only present in the insoluble fraction. The activity of the thioredoxin fusion protein containing functional Bmr2 protein closely matched the activity of the 6× His-tagged Bmr2 protein (Tables 1 and 2 and Figure S8), including lack of activity against sinapate. In contrast, no 4CL activity was detected against 4-coumarate, ferulate, caffeate and cinnamate at three concentrations each when the two mutant versions were used in the enzyme assay (Table 2). Hence, other 4CL proteins are probably responsible for the residual activity in *bmr2* stalks. Indeed, close examination of the immunoblot in Figure 4(a) indicates the presence of at least two proteins in some of the *bmr2* extracts. Given the cross-reactivity of the polyclonal antibody against Bmr2, and the slight variations in molecular weight (Figure 4b), these proteins are potentially other 4CLs.

Table 1 Kinetic parameters of Bmr2 and related 4CLs expressed with a 6× His tag in *E. coli* for various phenylpropanoid substrates

	K_m (μM)	V_{max} (nmol sec^{-1} mg $protein^{-1}$)	V_{max}/K_m (sec^{-1} mg $protein^{-1}$)	K_{cat} (sec^{-1})[a]
Bmr2 (64.6 kDa)				
4-coumaric acid	14 ± 1.7	52 ± 1.9	3.8×10^{-3}	3.4
Ferulic acid	15 ± 1.3	33 ± 2.3	2.3×10^{-3}	2.2
Caffeic acid	39 ± 6.1	46 ± 4.2	1.2×10^{-3}	3.0
Cinnamic acid	NA	10 ± 1.1	NA	0.64
Sinapic acid	ND	ND	NA	NA
Sb07g007810 (66.2 kDa)				
4-coumaric acid	7.5 ± 0.3	56 ± 0.58	7.4×10^{-3}	3.7
Ferulic acid	14 ± 3	27 ± 0.87	1.9×10^{-3}	1.8
Caffeic acid	16 ± 1.2	28 ± 2.5	1.8×10^{-3}	1.8
Cinnamic acid	81 ± 7	32 ± 2	0.40×10^{-3}	2.1
Sinapic acid	ND	ND	NA	NA
Sb10g026130 (64.3 kDa)				
4-coumaric acid	14 ± 1.0	17 ± 1.1	1.2×10^{-3}	1.1
Ferulic acid	X	X	NA	NA
Caffeic acid	36 ± 9.1	8.9 ± 1.0	0.25×10^{-3}	0.58
Cinnamic acid	134 ± 15	4.3 ± 0.24	0.03×10^{-3}	0.28
Sinapic acid	ND	ND	NA	NA
Sb04g031010 (65.1 kDa)				
4-coumaric acid	4.1 ± 0.17	23 ± 0.53	5.5×10^{-3}	1.5
Ferulic acid	X	X	NA	NA
Caffeic acid	78 ± 30	18 ± 5.2	0.22×10^{-3}	1.2
Cinnamic acid	273 ± 51	19 ± 3.4	0.07×10^{-3}	1.3
Sinapic acid	ND	ND	NA	NA

[a]K_{cat} was calculated using the molecular weights of the recombinant proteins as indicated in parentheses next to the protein names.
X, unable to determine due to a very narrow concentration range in which activity was observed; ND, not detectable; NA, not applicable.

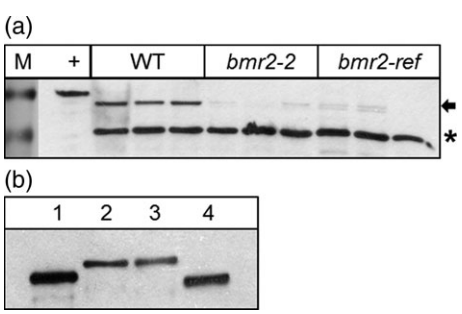

Figure 4. Western blot analysis of 4CL proteins.
(a) Immunoblot of protein extracts from wild-type and *bmr2* stalks. +, recombinant Bmr2 protein. Arrow, 4CL protein. Asterisk, *S*-adenosylmethionine synthetase. Three plants per genotype were assayed. M, broad-range pre-stained size markers (161-0318; Bio-Rad) showing BSA (80 kDa, top) and ovalbumin (49.1 kDa). The recombinant Bmr2 protein is slightly larger than the endogenous protein due to the presence of the 6× His tag used for purification.
(b) Cross-reactivity of the polyclonal antibody against Bmr2. Western blot based on 0.0825 μg recombinant 4CL protein per lane. Lane 1, Bmr2. Lane 2, Sb04g031010. Lane 3, Sb07g007810. Lane 4, Sb10g026130.

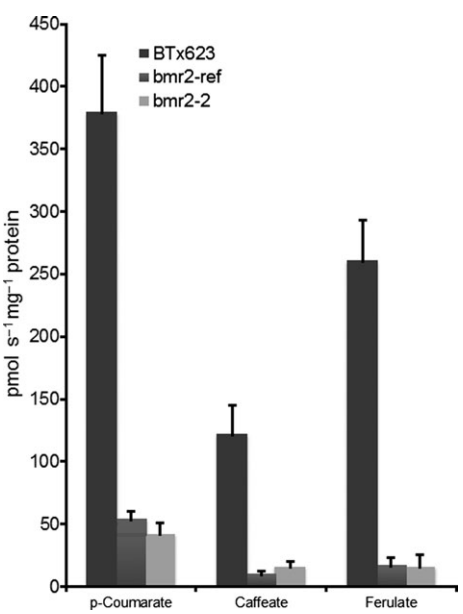

Figure 5. 4CL activity in 6-week-old stalk extracts from wild-type and *bmr2* mutants. Values are means ± SE (n = 3).

Structure of Bmr2 and closely related 4CL proteins

A 3D structural model for Bmr2 was built to better understand the impact of the two *bmr2* mutations. Building a meaningful structural model was possible, because Bmr2 shares a high level of sequence similarity to *Populus tomentosa* 4CL1 (Pto4CL1) (64% identity, 81% similarity), whose crystal structure has been determined (Hu *et al.*, 2010). None of the amino acid substitutions between Bmr2 and Pto4CL1 resulted in any major high-energy van der Waals contacts. Therefore, their overall structures were very

Table 2 4CL activity of Bmr2 and and the two mutant versions G111D and G262R expressed as thioredoxin fusion proteins in *Escherichia coli* for various phenylpropanoid substrates at three concentrations

		Velocity (nmol sec^{-1} mg $protein^{-1}$)		
Enzyme	Substrate	25 μM	50 μM	100 μM
Wild-type	4-coumarate	33	35	34
G111D		ND	ND	ND
G262R		ND	ND	ND
Wild-type	Ferulate	17	19	18
G111D		ND	ND	ND
G262R		ND	ND	ND
Wild-type	Caffeate	10	18	13
G111D		ND	ND	ND
G262R		ND	ND	ND
Wild-type	Cinnamate	4	4	6
G111D		ND	ND	ND
G262R		ND	ND	ND
Wild-type	Sinapate	ND	ND	ND
G111D		ND	ND	ND
G262R		ND	ND	ND

similar to each other, with a root mean square deviation of approximately 0.3 Å among the Cα atoms. In addition, the predicted secondary structural patterns of these 4CLs were almost identical. The aligned sequence of Bmr2 shows only one gap (Pro171 of Pto4CL1), which is located in the long loop connecting α6 and β7 (Figure S9). This high level of similarity made it possible to associate predicted structural differences with substrate specificity and kinetic parameters.

Most of the participating residues for both affinity for hydroxycinnamate-AMP and catalysis were almost identical with those of Pto4CL1. Specifically, the residues constituting the AMP binding motif were conserved (Figure S9). In addition, most of the residues constituting the binding pocket for the 4-hydroxyphenyl moiety, such as Tyr253, Ser257, Tyr347, Gly322, Gly348, Pro355 and Val356, were conserved. However, the superimposed active sites of Bmr2 and Pto4CL1 showed two differences: SbMet320 → PtoLys303 and SbAla323 → PtoGly306 (Figure 6). Similar changes have been observed among 4CL1 enzymes from various species, including *Physcomitrella patens*. The side chain of Lys303 in Pto4CL1 was indicated to play a role in positioning the 4-hydroxyl group through hydrogen bonding (Hu *et al.*, 2010), but because of the 3.52 Å distance between these groups, its role is thought to be relatively minor. Instead, the conserved neighboring Ser257 (Ser240 in Pto4CL1) forms a stable hydrogen bond with the 4-hydroxyl group (Figure 7b). As the substrate specificity of 4CL1s is thought be governed by a combination of hydrophobic interactions and the volume of the substrate binding pocket, the two substitutions in Bmr2 relative to Pto4CL1 (SbMet320 → PtoLys303 and SbAla323 → PtoGly306) may be responsible for differences in substrate

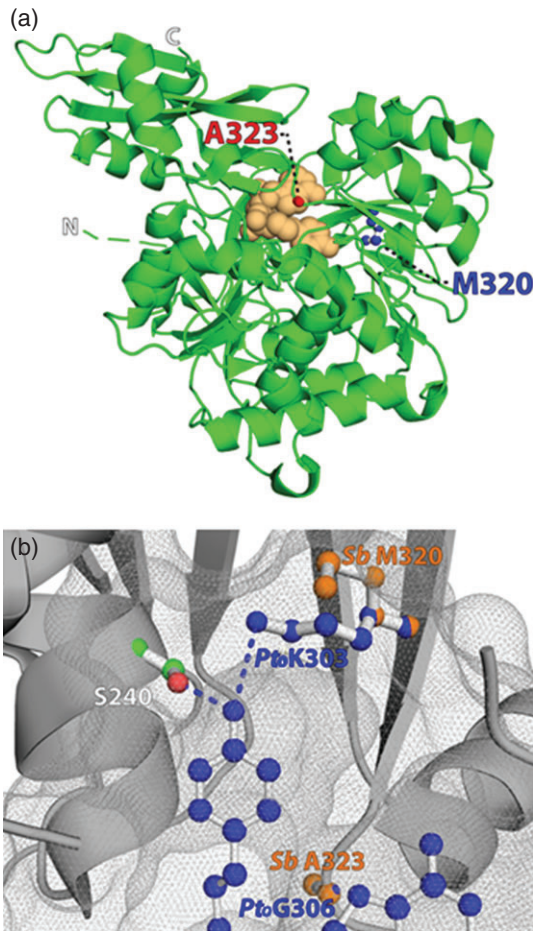

Figure 6. 3D model of Bmr2.
(a) Ribbon diagram representation of Bmr2 highlighting important amino acid differences compared with Pto4CL1.
(b) Structural differences in the active site pocket, shown as ball and stick. Residues from Bmr2 and Pto4CL1 are shown in orange and blue, respectively.

specificity and enzyme kinetics as a result of altered hydrophobic interactions (Hu *et al.*, 2010) (Table 1). Due to its high reliability the 3D model of Bmr2 can effectively serve as a foundation to explain how the amino acid changes caused by the *bmr2-ref* and *bmr2-2* mutations resulted in recombinant fusion proteins with no activity and greatly reduced 4CL activity *in planta*. The changes resulting from these mutations occur within predicted α-helices α3 and α9, respectively. Inserting charged side chains into a glycine position is expected to seriously affect both local conformation and protein folding, especially as all classified 4CL enzymes contain either glycine (85%) or alanine (15%) at position 111, whereas alanine is the most common amino acid at position 262 (85%), although At4CL2, At4CL3 and Pt4CL2 contain serine (Table S3). Impaired protein folding is consistent with the greatly reduced activity levels observed in the mutants (Figure 5) and the insolubility of mutant proteins expressed in *E. coli*. Specifically, Gly111Asp is expected to disrupt the hydrophobic interface between the perpendicularly oriented

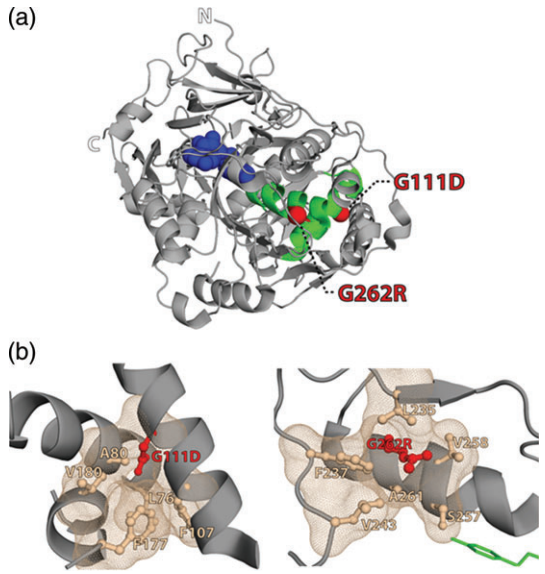

Figure 7. Structural representation of the effects of the *bmr2* mutations.
(a) Residues altered by the mutations are shown as red spheres, the secondary structures expected to be affected are shown in green, and the substrate analog is shown in blue.
(b) Hydrophobic pocket surrounding Gly111Asp (red) and Gly262Arg (red). Images were generated using open-source PyMOL™ (version 1.2; http://pymol.org/pymol).

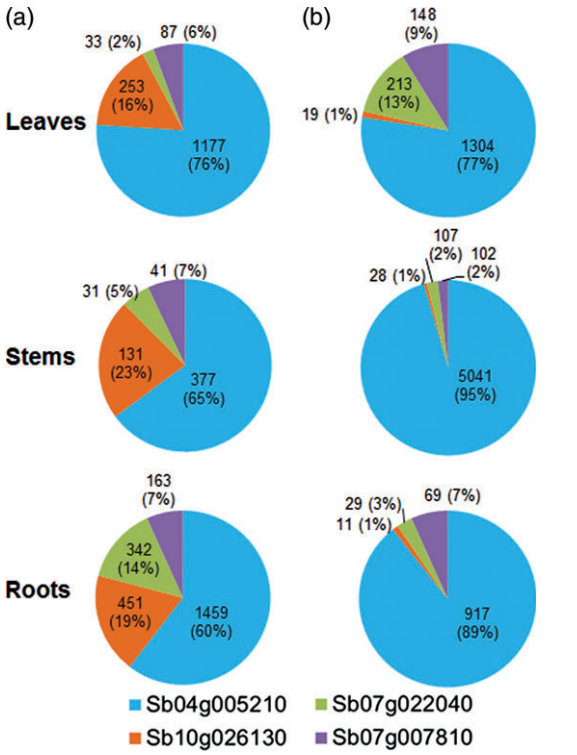

Figure 8. Relative contribution of each class I *4CL* gene to the total transcript pool. (a) Seedlings. (b) Pre-flowering plants. Numbers reflect template cDNA molecules present at the start of PCR.

long α-helices α2 and α3 comprising Ala80, Leu84, Val180 and Phe177. Gly262Arg is also expected to disrupt the hydrophobic interface between α9 and the β-sheets β9, β11 and β12 comprising Leu235, Phe237, Val243 and Ile291. The Gly262Arg substitution may also compromise the hydrogen bonding between nearby Ser257 and the 4-hydroxyl group of the substrate (Figure 7b). Therefore, even if the folding integrity of this mutant version of Bmr2 were maintained *in planta*, the resulting protein would be expected to have altered affinity for hydroxycinnamate-AMP.

Gene expression of *Bmr2*, closely related sorghum *4CL* genes, and other genes involved in monolignol biosynthesis

Quantitative RT-PCR was performed to compare *Bmr2* gene expression with the other sorghum class I *4CL* genes. The *Bmr2* gene was the most highly expressed of the bona fide *4CL* genes in all organs and developmental stages in the BTx623 background (Figures 8 and 9), consistent with the results of Western blotting (Figure 4a). Comparison of transcript numbers between the seedling and pre-flowering stages indicated that the proportion of *Bmr2* transcripts among the total number of class I *4CL* transcripts increased as the plants developed (Figure 8). However, comparing gene expression in the leaves between the two stages, the number of *Bmr2* transcripts was not statistically significantly different. This probably reflects the fact that these organs were developmentally similar at both stages (Figure 8). Instead, the main difference was in the relative contribution of the *4CL* paralogs. In contrast, when comparing seedling stems and pre-flowering upper stems, there was a ninefold increase in *Bmr2* expression in the pre-flowering stems, which reflects the considerable increase in cell-wall lignification between these two development stages. In the roots, the total number of transcripts from the class I *4CL* genes decreased with age. The number of transcripts from genes Sb07g022040 and Sb07g007810 remained unchanged or increased modestly over the course of development, but expression of gene Sb10g026130 was more variable (Figure 8 and Table S4). In seedlings, class I *4CL* genes were expressed more abundantly in the roots, but the same genes were expressed more abundantly in the stem in pre-flowering plants, which reflects the difference in the main site of lignification between these stages (Figure 8). Sb04g031010, which belongs to class II, was primarily expressed in aerial parts of the plant (Table S4B).

The effects of the *bmr2* mutations on *4CL* expression in these organs and at these two stages were also examined (Figure 9). The biggest impact of the *bmr2-2* mutation was on the abundance of the *bmr2* transcript itself, evidenced by increases in expression levels of 2.5-, 5- and 8.5-fold in seedling leaves, stems and roots, respectively. In pre-flowering *bmr2* plants, expression of *Bmr2* was up-regulated

13.8- and 2.8-fold in the leaves and roots, respectively, but no statistically significant difference was detected in the upper stems at this stage. To test whether the observed up-regulation of the *Bmr2* gene in the mutant plants was background- and/or allele-dependent, we examined its expression in the leaves of five mutant and five phenotypically wild-type F_3 sibling plants derived from the AMP11 × Theis cross and segregating for the *bmr2-ref* allele. Although more variability is expected when comparing F_3 siblings rather than isolines due to differences in genetic background and rates of development, statistically significant differences were detected between the mutant and wild-type siblings, with a mean up-regulation of 1.6-fold (Table S4). Given the predominant role of Bmr2 during the course of development (Figure 9 and Table S4), *Bmr2* expression was also compared between upper (younger) and lower (older) internodes in pre-flowering wild-type and *bmr2-ref* plants. Consistent with the initial expression data, *Bmr2* expression was lower in the upper (younger) internodes of both wild-type and *bmr2* plants (Table S4C), and the gene expression levels among these genotypes were not statistically significantly different (Table S4D). *Bmr2* expression in the lower (older) internodes of wild-type and *bmr2* plants was eight- and 12-fold higher, respectively, than in the upper (younger) internodes, and the difference in expression between wild-type and *bmr2* mutants was statistically significant. The effect of the mutant background on the other *4CL* genes investigated was more variable, but a general trend toward up-regulated expression of these genes was observed (Figure 9 amd Table S4).

In order to assess whether the loss of Bmr2 activity also affected the gene expression levels of other monolignol-related genes, the expression levels of the genes encoding cinnamate 4-hydroxylase [C4H; two paralogs, one implicated in lignification (*C4H1*) and one in defense responses (*C4H2*)], *p*-coumaroyl ester 3′ hydroxylase (C3′H), cinnamoyl CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD) and caffeic acid *O*-methyl transferase (COMT) were compared among wild-type and *bmr2* plants. Modest but statistically significant increases in expression in the *bmr2* mutant were observed for *CCR* and the *C4H* paralog implicated in lignification (*C4H1*), but not for the other genes (Table S4E).

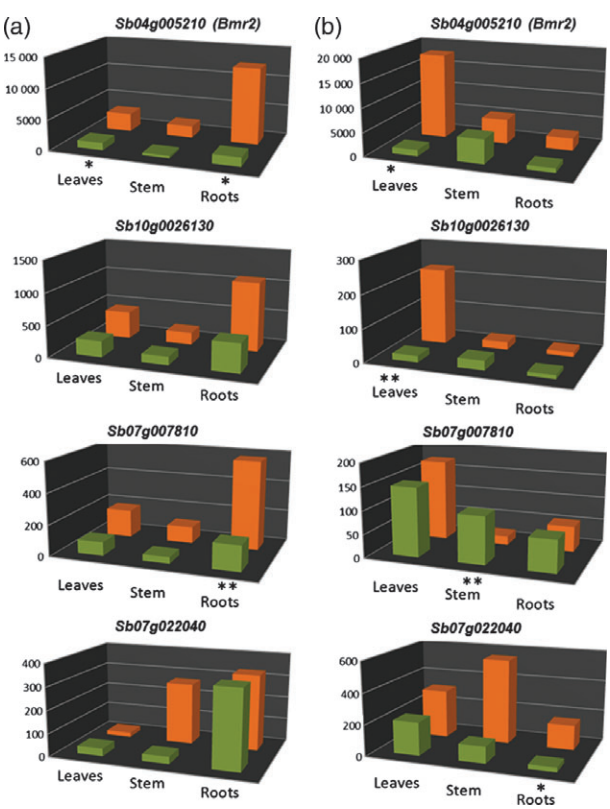

Figure 9. Effect of the *bmr2-2* mutation on expression of genes encoding class I 4CL proteins.
(a) Gene expression in seedlings.
(b) Gene expression in pre-flowering plants. Green bars, wild-type plants. Orange bars, *bmr2-2* plants. Statistically significant differences are indicated by asterisks: $*P < 0.1$, $**P < 0.05$. Vertical scale: number of template cDNA molecules present at the start of PCR.

Identification of motifs in the promoters of bona fide sorghum *4CL* genes

Given the increased expression of some of the *Bmr2* paralogs in the *bmr2* mutant, an *in silico* promoter analysis was performed. The presence of three sequence motifs [P-box, L-box (AC elements) and A-box] is considered a hallmark of the promoter regions of *4CL* and *phenylalanine ammonia lyase* (*PAL*) genes involved in lignification (Logemann *et al.*, 1995; Chen *et al.*, 2000; Stracke *et al.*, 2001). Of the five sorghum genes identified as belonging to the bona fide 4CL clade, only *Bmr2* contained all three motifs in its promoter region (Figure 10). In addition to these motifs, several putative regulatory sequences that are responsive to light and plant hormones were identified (Table S4). Of the remaining sorghum genes in the *4CL* clade, only Sb10g026130 and Sb07g022040 contained AC elements in their promoter regions, at positions −602 and −2880, respectively, together with several other regulatory elements (Table S5). However, their position is much more distal than observed for the AC elements of reported bona fide *4CL* genes, which are usually located within 400 bp upstream of the transcription start site. These findings, together with the fact that the *Sb07g007810* promoter lacks AC elements, indicate that additional regulatory sequences must exist to confer the moderate levels of expression of other *4CL* genes observed in lignifying tissues.

DISCUSSION

Both *bmr2* mutant alleles contain missense mutations, resulting in incorporation of amino acids of different character, which plausibly prevent proper folding of the proteins based

Figure 10. Promoter structure of *Bmr2* displaying the conserved motifs P-box, L-box and A-box identified in *4CL* genes from other species. A G-box, the target for an ABA-regulated bZIP transcription factor, is also present. Grey shading marks TATA-box.

on the molecular modeling. This is consistent with the immunoblot analysis of stem extracts and the loss of solubility of the recombinant mutant proteins expressed in *E. coli*. Despite the severe effect of the mutations on Bmr2, with mutant stems displaying only approximately 13% of wild-type activity, the impact on cell-wall composition is modest overall, with a 20% reduction in lignin in the stem and leaves and no obvious detrimental effects on plant growth and development. Tracking *Bmr2* expression over the course of development, combined with histochemical data, suggests that the importance of Bmr2 increases over time, but not until after formation of proto-xylem in the stem. This limits the effect of the *bmr2* mutation on stem xylem integrity. Our data indicate that lack of Bmr2 is compensated for by higher expression of several paralogs, and that the enzymes they encode are functional 4CLs, some of which have activity very similar to Bmr2. There is also evidence for both tissue-specific and developmental differences in the way 4CL activity is provided. For example, *Sb07g007810* transcripts are proportionally more abundant in the leaves of pre-flowering plants than in stems (Figures 8 and 9). The activity towards ferulate of the enzyme encoded by this gene may explain the accumulation of ferulate in stem tissue but not in leaf tissue of the *bmr2* mutant. Ferulate accumulation in *bmr2* stem extracts is unexpected as this compound is generally not considered an intermediate in the monolignol biosynthetic pathway (Nair *et al.*, 2004). However, there is evidence that ferulate can be synthesized from 4-coumarate, especially when the normal biosynthetic route is blocked (Meyermans *et al.*, 2000).

A major benefit of using a *4CL* mutant rather than a transgenically down-regulated line is that changes in the expression of the mutated gene can be investigated. Unlike nonsense mutations resulting in nonsense-mediated mRNA decay (Rebbapragada and Lykke-Andersen, 2009), missense mutations do not automatically affect transcript levels. The fact that *Bmr2* transcripts are more abundant in the mutants is direct evidence for an auto-regulatory mechanism that either involves enhanced gene expression or an increased half-life of the transcript. Increased transcript levels resulting from a missense mutation in the *SHOOT MERISTEMLESS* (*STM*) gene were reported by Takano *et al.* (2010). STM is a DNA-binding protein, so a mutation in this gene may feasibly affect transcription directly. In the case of 4CL, a direct effect on transcription is not as likely, and identifying the nature of this feedback mechanism will require additional experiments aimed at establishing whether the reduced amount of protein and/or accumulation of substrate are the primary trigger for the increased transcript levels. Several enzymes involved in phenylpropanoid biosynthesis have been hypothesized to be part of multi-enzyme complexes (Czichi and Kindl, 1977; Burbulis and Winkel-Shirley, 1999; Achnine *et al.*, 2004) that channel precursors from one enzyme to the next to ensure synthesis of various compounds from the same set of precursors. If such enzyme complexes involving 4CL exist in sorghum, a shortage of 4CL may trigger a response leading to production of more enzyme via a complex-sensing mechanism. This may also explain the higher expression levels of several other genes involved in monolignol biosynthesis, specifically *C4H1* and *CCR*, which are responsible for generating the 4CL substrate 4-coumarate and for converting the 4CL product 4-coumaroyl CoA to 4-coumaraldehyde, respectively. The lack of statistically significant changes in expression of the defense-related *C4H2* gene and genes encoding enzymes that act further downstream (COMT and CAD) suggests that these genes are under a different control mechanism. Based on these expression data, a global expression profiling experiment may prove of interest to investigate how the *bmr2* mutations affect the expression of other genes involved in the metabolism of phenolic compounds. Such an analysis may reveal groups of genes under common regulatory control, as has been observed in other species (Ehlting *et al.*, 2005). Ultimately, the three class I *4CL* paralogs are up-regulated in some tissues, and the proteins they encode are probably responsible for the observed 4CL activity in the *bmr2* mutants, thereby enabling monolignol biosynthesis. The weak signal on the immunoblot of the *bmr2* stem extracts suggests that even small amounts of these 4CLs are adequate.

The variation in cell-wall architecture and cell-wall composition between grasses and dicots is thought to result from changes in the roles that specific classes of enzymes play in cell-wall biogenesis. As substrate preference and catalytic activity of sorghum Bmr2 appear similar overall to what has been reported for dicots, and given the presence of two classes of *4CL* genes that mirror those of dicots, the differences in the role that 4CL plays in cell-wall biosynthesis in sorghum versus angiosperm dicots are manifested as spatial and temporal variation in the expression of *4CL* gene family members. The most compelling evidence for differences is that, in Arabidopsis, two *4CL* genes (*4CL1* and *4CL2*) are actively involved in the formation of monolignols (Ehlting *et al.*, 1999). Goujon *et al.* (2003) speculated that mutations in both of these genes are required to create a

mutant phenotype consistent with reduced 4CL activity. In sorghum, based on the *bmr2* phenotype and gene expression data, Bmr2 is clearly the predominant 4CL, both in seedlings and older plants, and in all tissues that were investigated. Nonetheless, the presence of several *4CL* paralogs, some of which are up-regulated in response to the *bmr2* mutation, indicates a level of functional redundancy that reflects the key role of 4CL in plant metabolism, located at the intersection of flavonoid, monolignol and hydroxycinnamic acid biosynthesis.

EXPERIMENTAL PROCEDURES

Source of the *bmr2* alleles

The source of the *bmr2-ref* allele (Porter *et al.*, 1978; Saballos *et al.*, 2008) was line AMP11 (in the OK11 background; National Plant Germplasm System PI number 602899), and the *bmr2-2* allele originated from line 562 of the sorghum TILLING population (Xin *et al.*, 2008). Comparisons were made with their corresponding wild-type controls OK11 and BTx623.

Mapping population

Cultivar 'Theis' was used as pollen donor for the male-sterile line AMP11 to generate 200 F_2 lines for mapping. Twenty additional F_3 mapping lines were obtained from heterozygous F_2 plants.

Mapping of the *bmr2* locus

Fifteen SSR markers developed by Bhattramakki *et al.* (2000) and found to be polymorphic between the parents of the F_2 population were used to determine an initial map location for *bmr2*. PCR primers flanking putative SSRs within a 4 Mb region including the *bmr2* locus were designed using the SSR Extractor (http://www.aridolan.com/ssr/ssr.aspx). Uniformly spaced 2 kb genomic fragments in the 2.5 Mb region were cloned and sequenced in order to identify SNPs, which were used for fine mapping (Table S1).

Identification of *bmr2* mutations

The *Bmr2* gene was sequenced as three overlapping fragments (−678 to 1649 bp, 1445–3528 bp and 3459–4465 bp, Table S1) generated by PCR, and subcloned in pT7-Blue3 (EMD Biosciences; http://www.emdbiosciences.com). Sequence analysis was performed using Biology WorkBench 3.2 (http://workbench.sdsc.edu/).

Phylogenetic analysis

Multiple sequence alignments were generated using CLUSTAL W version 3.2 (Thompson *et al.*, 1994) with the BLOSUM matrix and the 'accurate' default parameters (http://workbench.sdsc.edu; Dayhoff and Orcutt, 1979), and used for a phylogenetic analysis using MrBayes software (http://mrbayes.csit.fsu.edu/index.php) with the priors for the amino acid model set to 'mixed' to allow jumping between the fixed models (Huelsenbeck and Ronquist, 2005). The model-jumping option was selected as no prior knowledge of the best evolutionary model for the acyl CoA synthetase gene family was assumed. Bayesian Markov chain Monte Carlo (MCMC) model jumping is an alternative to model selection prior to the analysis. Two independent runs of MCMC analysis were performed for 1 125 000 generations, with sampling every 100 generations until the standard deviation of the split frequencies was stable. The Output.con file was read into iTOL (http://itol.embl.de) for visualization and analysis (Letunic and Bork, 2006, 2011).

Cloning and heterologous expression of *Bmr2* cDNA

The *Bmr2* coding region was cloned from the sorghum EST GABR1_8_D10.g1_A002 (GenBank accession CX613521) into the *Kpn*I and *Not*I restriction sites of expression vector pET30a (EMD Biosciences), and sequenced to confirm accuracy. The plasmid was introduced into Rosetta R2 *E. coli* cells for protein expression. Cultures inoculated from a single colony were grown to log phase at 37°C, transferred to 20°C, and induced to produce protein for approximately 18 h following addition of 0.1 mm isopropyl β-D-1-thiogalactopyranoside. Soluble protein was extracted by sonication. The expressed protein contained an N-terminal 6× His tag, and was captured on a nickel resin column and eluted using imidazole. Induction of the expressed protein and protein purification were monitored by SDS–PAGE. Expression vectors containing the paralogous cDNAs Sb10g0026130 (CN143608), Sb07g007810 (CD429206), Sb04g031010 (CX609752) and Sb07g022040 (CN129170) were prepared similarly, but soluble protein could only be obtained from the first three. Wild-type Bmr2 protein and the two mutant versions (Gly111Asp and Gly262Arg) were expressed as thioredoxin fusion proteins using vector pET32a (EMD Biosciences) to enhance the solubility of the mutant versions, and purified as described above.

Extraction of proteins from sorghum stalks

Eight-week-old stalks from greenhouse-grown plants were ground to a fine powder under liquid nitrogen. Three volumes of cold extraction buffer containing 100 mM Tris, pH 7.8, 5 mM $MgCl_2$, 5 mM DTT, 10 μl per ml sample volume of protease inhibitor P9599 (Sigma, http://www.sigmaaldrich.com/), 15% v/v ethylene glycol and 5% w/v polyvinylpyrrolidone were added to the ground tissue. After mixing the samples at 4°C using a rotary mixer for 1 h, and centrifugation at 18 620 ***g*** for 15 min, the supernatant was collected. The protein content of the extracts was determined using the Pierce 660 nm protein assay (http://www.piercenet.com) with lysozyme as a protein standard.

Immunoblotting

Purified recombinant 4CL proteins or proteins in the plant extracts were separated on a 12% SDS gel and transferred to a nitrocellulose membrane in a wet blot system for 45 min at 65 V in 10 mM Tris, 100 mM glycine and 8% v/v methanol. Following transfer, the membrane was stained with Ponceau S to verify transfer, and blocked using 3% non-fat dry milk in Tris-buffered saline (TBS) containing 0.5% v/v Tween-20 (TBST) for 1 h. The membrane was probed using primary antibodies [polyclonal rabbit antisera against recombinant Bmr2 and *S*-adenosylmethionine synthetase (Cocalico Biologicals http://www.cocalicobiologicals.com)] diluted 1:5000 and 1:10 000, respectively, for 1 h, followed by three 5 min washes with TBST. The secondary antibody, goat anti-rabbit IgG+horseradish peroxidase (A-0545; Sigma Aldrich), was diluted 1:10 000, added to the membrane, and incubated for 1 h. The membrane was washed twice in TBST for 5 min and once in TBS + 0.5 M NaCl for 5 min. Secondary antibody was detected by chemiluminescence using Amersham ECL Western blotting reagent (http://www.gelifesciences.com).

4-coumarate CoA ligase enzyme kinetics

The continuous spectrophotometric assay for 4CL activity and the reaction conditions were as previously described (Knobloch and Hahlbrock, 1977) with minor changes. The reaction was performed in 100 mM Tris/HCl, pH 7.8, 5 mM $MgCl_2$, 5 mM ATP and 0.3 mM coenzyme A at 25°C. Substrate concentrations were varied from 5 to 150 μM for the substrates *p*-coumarate, cinnamate, caffeate, ferulate

and sinapate. The reactions were initiated by adding 30 μl of the plant protein extract (300 μg/ml total protein) or 10 μl rmbinant protein diluted to 6 ng/μl. The change of absorbance was monitored over 10 sec intervals for 5 min at the reported absorption maxima of the phenolic CoA products (Lee *et al.*, 1997; Rautengarten *et al.*, 2010) using a 96-well plate reader. Samples that contained no enzyme or no extract lacked activity. Four replicates were performed for each assay. Absorbance data were analyzed using KaleidaGraph 4.0 (Synergy Software, http://www.synergy.com)) and SAS 9.1 (SAS Institute, http://www.sas.com/). The rate of production of phenolic CoA molecules was calculated based on a simple two-parameter regression line fitted through the absorbance data using a linear least-squares model. The kinetic parameters K_m and V_{max} were calculated using a non-linear least squares method under the 'general curve fit' procedure in KaleidaGraph 4.0. In cases where substrate inhibition was evident from a decline in velocity with increasing substrate concentration, V_{max} and K_m were calculated from the intercept and slope, respectively, of the line tangent to the curve in a Lineweaver–Burk plot (1/*V* versus 1/[substrate]) (Schulz, 1994).

Structural modeling of Bmr2 and two mutant versions

The crystal structure of *Populus tomentosa* 4CL1 bound to APP (an analog of adenosine 5′-coumaroyl phosphate) (Protein DataBank ID 3NI2) (Hu *et al.*, 2010) was used to build a molecular model for Bmr2. The amino acid substitutions were performed starting with the spatial coordinates of Pto4CL1, and regularized using defined geometric constraints for bonds, angles, planes, non-bonded contacts and torsion constraints in Coot (Emsley *et al.*, 2010). After geometric regularization, quick global energy minimization was performed using CNS version 1.1 (Brunger *et al.*, 1998), which uses the potential function parameters of CHARMM19 (http://www.charmm.org). The substrate position was generated using the solid docking module of QUANTA (http://www.accelrys.com), which is based on conformational space, followed by energy minimization using CNS version 1.1.

Gene expression analysis of sorghum *4CL* genes

Tissue from developing leaves, roots and stems was harvested from greenhouse-grown BTx623 and *bmr2-2* plants (two each) at stage 2 (seedlings at the five-leaf stage) and stage 4 (pre-flowering with flag leaf visible). At stage 2, all the leaf tissue and the entire root ball were harvested. The stem sample consisted of the stacked internodes including the terminal meristem after removal of the leaf sheaths. At stage 4, the whorl of expanding leaves, approximately 10 cm long root tips, and the first internode below the peduncle of the immature developing panicle were collected. In addition, leaf tissue from the whorl of five wild-type and five *bmr2-ref* F_3 plants derived from the mapping population was collected at approximately 35–50 days after planting, depending on when the *bmr* phenotype became visible. Furthermore, an expression analysis was performed on OK11 and *bmr2-ref* plants. The stems of three stage 4 plants from both genotypes were collected and dissected into upper internodes (first two or three internodes below the developing panicle) and lower internodes (first two or three internodes above the soil line). All tissue samples were frozen and ground in liquid nitrogen and kept at –80°C until extraction of RNA from 100 mg tissue using a MasterPure Plant RNA purification kit (Epicenter; http://epibio.com). Total cDNA was synthesized using an iScript III cDNA synthesis kit (Bio-Rad, http://www.bio-rad.com/).

Quantitative RT-PCR primers were designed using Integrated DNA Technologies software (https://www.idtdna.com/Scitools/Applications/RealTimePCR), spanning introns to enable detection of genomic DNA contamination (Table S1). Sorghum GAPD and ubiquitin were used to normalize cDNA abundance (Table S1). To test the specificity of the expression primers, cDNA from line BTx623 was used for amplification of the target genes. Primer specificity was verified by cloning and sequencing PCR products obtained from BTx623 cDNA. In addition, melting curve analyses were performed for each quantitative RT-PCR reaction to verify single-amplicon amplification.

The amplification efficiency of each primer set was determined using purified plasmids containing each of the target fragments linearized with *Sma*I and diluted from 10^8 to 10 molecules of plasmid per microliter. Yeast tRNA was added to avoid degradation of diluted samples. Quantitative RT-PCR at each concentration for each plasmid–primer combination was performed in triplicate using IQ™ Sybr® Green Supermix real-time PCR mix (Bio-Rad). The threshold cycle (C_t) and $\log_{10}$ of the template concentration were plotted to calculate the standard curve and efficiency of amplification. The equations for the regression lines between template concentration and C_T for each primer pair were used to calculate transcript numbers in the samples. The $\Delta\Delta C_t$ method described by Schmittgen (2006) was used for relative quantification to determine the fold difference between the expression of each gene in wild-type and mutant tissue. For this analysis, the *GAPD* gene was used as the reference gene.

Gene expression of sorghum monolignol biosynthetic genes

Quantitative RT-PCR using the $\Delta\Delta C_t$ method as described above was used to compare gene expression in the lower internodes of pre-flowering wild-type and *bmr2-ref* plants for the following monolignol biosynthetic genes: *C4H1* (cinnamate 4-hydroxylase; Sb02g010910, the closest homolog to Arabidopsis *C4H*), *C4H2* (Sb03g038160; implicated in the defense response), *C3′*H (*p*-coumaroyl ester 3′-hydroxylase; Sb09g024210), *CCR* (cinnamoyl CoA reductase; Sb07g021680), *CAD2* (cinnamyl alcohol dehydrogenase; Sb04g005950; Saballos *et al.*, 2009) and *COMT* (caffeic acid *O*-methyl transferase; Sb07g003860; Bout and Vermerris, 2003). Primer sequences are listed in Table S1. The analysis was performed on the same cDNA samples that were used to quantify differences in *4CL* expression.

In silico promoter analysis of the sorghum *4CL* family members

The sequence of the complete intergenic region upstream of genes Sb04g005210, Sb10g026130, Sb07022040, Sb07g007810 and Sb04g031010 was downloaded from the sorghum genome database (http://www.phytozome.net/sorghum). The programs NSITE version 2.2004 and TSSP (Softberry Inc., http://linux1.softberry.com/berry.phtml) were used to query the sequences for regions homologous to known promoter elements from other species.

ACKNOWLEDGEMENTS

We thank Tammy Griess and Dr Aaron Saathoff (USDA-ARS, Lincoln, NE) for technical assistance with the biochemical experiments, Dr Stuart McDaniel (Biology department, University of Florida) for helpful suggestions on the phylogenetic analysis, Drs Marty Cohn and Brad Barbazuk (University of Florida Genetics Institute) for use of their thermal cycler, and Dr Gary Peter (University of Florida Genetics Institute) for use of his microscope. This research was supported by the Office of Science (Biological and Environmental Research), US Department of Energy (grant DE-FG02-07ER64458 to W.V., S.E.S, and J.P.), and the US Department of Agriculture (grant 35318-17454 to C.H.K.), with additional funding from the University of Florida.

SUPPPRTING INFORMATION

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

Supplement Methods

Figure S1. Enzymatic reaction catalyzed by 4CL.

Figure S2. The *bmr2* phenotype.

Figure S3. Histochemical analysis of *bmr2*.

Figure S4. Chemical composition of wild-type and *bmr2* midribs.

Figure S5. 4-coumarate and ferulate in wild-type and *bmr2* midribs.

Figure S6. Chemical composition of wild-type and *bmr2* stover.

Figure S7. 4-coumarate and ferulate in tissue extracts.

Figure S8. Kinetic parameters of Bmr2 and related 4CLs with various phenylpropanoid substrates.

Figure S9. Sequence alignment of Bmr2 and Pto4CL1.

Table S1. Sequences of PCR primers.

Table S2. 4CL and 4CL-like sequences included in the phylogenetic analysis.

Table S3. Alignment of 4CL sequences.

Table S4. Statistical analysis of quantitative RT-PCR data.

Table S5. *In silico*-identified promoter elements of *Bmr2*, *Sb10g026130* and *Sb07g022040*.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Achnine, L., Blancaflor, E.B., Rasmussen, S. and Dixon, R.A. (2004) Colocalization of l-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell*, **16**, 3098–3109.

Allina, S.M., Pri-Hadash, A., Theilmann, D.A., Ellis, B.E. and Douglas, C.J. (1998) 4-coumarate:coenzyme A ligase in hybrid poplar: properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. *Plant Physiol.* **116**, 743–754.

Bhattramakki, D., Dong, J.M., Chhabra, A.K. and Hart, G.E. (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. *Genome*, **43**, 988–1002.

Bout, S. and Vermerris, W. (2003) A candidate-gene approach to clone the sorghum *Brown midrib* gene encoding caffeic acid *O*-methyltransferase. *Mol. Gen. Genomics.* **269**, 205–214.

Brunger, A.T., Adams, P.D., Clore, G.M. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

Burbulis, I.E. and Winkel-Shirley, B. (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. *Proc. Natl Acad. Sci. USA*, **96**, 12929–12934.

Chabannes, M., Barakate, A., Lapierre, C., Marita, J.M., Ralph, J., Pean, M., Danoun, S., Halpin, C., Grima-Pettenati, J. and Boudet, A.M. (2001) Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *Plant J.* **28**, 257–270.

Chen, F. and Dixon, R.A. (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* **25**, 759–761.

Chen, C.Y., Meyermans, H., Burggraeve, B., De Rycke, R.M., Inoue, K., De Vleesschauwer, V., Steenackers, M., Van Montagu, M.C., Engler, G.J. and Boerjan, W.A. (2000) Cell-specific and conditional expression of caffeoyl-coenzyme A-3-*O*-methyltransferase in poplar. *Plant Physiol.* **123**, 853–867.

Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A. *et al.* (2005) Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate:CoA ligase network in Arabidopsis: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry*, **66**, 2072–2091.

Cukovic, D., Ehlting, J., VanZiffle, J.A. and Douglas, C.J. (2001) Structure and evolution of 4-coumarate:coenzyme A ligase (*4CL*) gene families. *Biol. Chem.* **382**, 645–654.

Czichi, U. and Kindl, H. (1977) Phenylalanine ammonia lyase and cinnamic acid hydroxylase as assembled consecutive enzymes on microsomal membranes of cucumber cotelydons. *Planta*, **134**, 133–143.

Dayhoff, M.O. and Orcutt, B.C. (1979) Methods for identifying proteins by using partial sequences. *Proc. Natl Acad. Sci. USA*, **76**, 2170–2174.

Ehlting, J., Buttner, D., Wang, Q., Douglas, C.J., Somssich, I.E. and Kombrink, E. (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* **19**, 9–20.

Ehlting, J., Shin, J.J.K. and Douglas, C.J. (2001) Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. *Plant J.* **27**, 455–465.

Ehlting, J., Mattheus, N., Aeschliman, D.S. *et al.* (2005) Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in phenylpropanoid metabolism and transcriptional regulators of fiber differentiation. *Plant J.* **42**, 618–640.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D*, **66**, 486–501.

Goujon, T., Sibout, R., Eudes, A., MacKay, J. and Jouanin, L. (2003) Genes involved in the biosynthesis of lignin precursors in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* **41**, 677–687.

Gui, J., Shen, S. and Li, L. (2011) Functional characterization of evolutionarily divergent 4-coumarate:coenzyme A ligases in rice. *Plant Physiol.* **157**, 574–586.

Guillaumie, S., San-Clemente, H., Deswarte, C., Martinez, Y., Lapierre, C., Murigneux, A., Barrière, Y., Pichon, M. and Goffner, D. (2007) MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. *Plant Physiol.* **143**, 339–363.

Hamberger, B. and Hahlbrock, K. (2004) The *4-coumarate:CoA ligase* gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc. Natl Acad. Sci. USA*, **101**, 2209–2214.

Hu, W.J., Kawaoka, A., Tsai, C.J., Lung, J.H., Osakabe, K., Ebinuma, H. and Chiang, V.L. (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). *Proc. Natl Acad. Sci. USA*, **95**, 5407–5412.

Hu, W.J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.J. and Chiang, V.L. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat. Biotechnol.* **17**, 808–812.

Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L., Li, D., Jiang, X.-N. and Wang, D.-C. (2010) Crystal structures of *a Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. *Plant Cell*, **22**, 3093–3104.

Huelsenbeck, J.P. and Ronquist, F. (2005) Bayesian analysis of molecular evolution using MrBayes. In Statistical Methods in Molecular Evolution (Nielsen R ed.) New York: Springer, pp. 183–232.

Jackson, L.A., Shadle, G.L., Zhou, R., Nakashima, J., Chen, F. and Dixon, R.A. (2008) Improving saccharification efficiency of alfalfa stems through modification of the terminal stages of monolignol biosynthesis. *Bioenergy Res.* **1**, 180–192.

Jones, L., Ennos, A.R. and Turner, S.R. (2001) Cloning and characterization of *irregular xylem4* (*irx4*): a severely lignin-deficient mutant of Arabidopsis. *Plant J.* **26**, 205–216.

Kajita, S., Hishiyama, S., Tomimura, Y., Katayama, Y. and Omori, S. (1997) Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate:coenzyme A ligase is depressed. *Plant Physiol.* **114**, 871–879.

Khurana, P., Gokhale, R.S. and Mohanty, D. (2010) Genome scale prediction of substrate specificity for acyl adenylate superfamily of enzymes based on active site residue profiles. *BMC Bioinformatics*, **11**, 57.

Knobloch, K.H. and Hahlbrock, K. (1975) Isoenzymes of *para*-coumarate:CoA ligase from cell-suspension cultures of *Glycine max*. *Eur. J. Biochem.* **52**, 311–320.

Knobloch, K.H. and Hahlbrock, K. (1977) 4-Coumarate:CoA ligase from cell-suspension cultures of *Petroselinum hortense* Hoffm: partial purification, substrate specificity and further properties. *Arch. Biochem. Biophys.* **184**, 237–248.

Lee, D. and Douglas, C.J. (1996) Two divergent members of a tobacco *4-coumarate:coenzyme A ligase* (*4CL*) gene family: cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *Plant Physiol.* **112**, 193–205.

Lee, D., Meyer, K., Chapple, C. and Douglas, C.J. (1997) Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. *Plant Cell*, **9**, 1985–1998.

Letunic, I. and Bork, P. (2006) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, **23**, 127–128.

Letunic, I. and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–W478.

Lindermayr, C., Mollers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H. and Ebel, J. (2002) Divergent members of a soybean (*Glycine max* L.) *4-coumarate:coenzyme A ligase* gene family: primary structures, catalytic properties, and differential expression. *Eur. J. Biochem.* **269**, 1304–1315.

Logemann, E., Parniske, M. and Hahlbrock, K. (1995) Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. *Proc. Natl Acad. Sci. USA*, **92**, 5905–5909.

Nair, R.B., Bastress, K.L., Rugger, M.O., Denault, J.W. and Chapple, C.C. (2004) The *Arabidopsis thaliana REDUCED EPIDERMAL FLUORESCENCE1* gene encodes and aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell*, **16**, 544–554.

Nakashima, J., Chen, F., Jackson, L., Shadle, G. and Dixon, R.A. (2008) Multi-site genetic modification of monolignol biosynthesis in alfalfa (*Medicago sativa*): effects on lignin composition in specific cell types. *New Phytol.* **179**, 738–750.

Penning, B.W., Hunter, C.T. III, Tayengwa, R. *et al.* (2009) Genetic resources for maize cell wall biology. *Plant Physiol.* **151**, 1703–1728.

Porter, K.S., Axtell, J.D., Lechtenberg, V.L. and Colenbrander, V.F. (1978) Phenotype, fiber composition and *in vitro* dry matter disappearance of chemically induced *brown midrib* (*bmr*) mutants of sorghum. *Crop Sci.* **18**, 205–208.

Ralph, J., Lundquist, K., Brunow, G. *et al.* (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochem. Rev.* **3**, 29–60.

Rautengarten, C., Baidoo, E., Keasling, J.D. and Scheller, H.V. (2010) A simple method for enzymatic synthesis of unlabeled and radiolabeled hydroxycinnamate-CoA. *Bioenergy Res.* **3**, 115–122.

Rebbapragada, I. and Lykke-Andersen, J. (2009) Execution of nonsense-mediated mRNA decay: what defines a substrate? *Curr. Opin. Cell Biol.* **21**, 394–402.

Saballos, A., Vermerris, W., Rivera, L. and Ejeta, G. (2008) Allelic association, chemical characterization and saccharification properties of *brown midrib* mutants of sorghum (*Sorghum bicolor* (L.) Moench). *Bioenergy Res.* **2**, 198–204.

Saballos, A., Ejeta, G., Sanchez, E., Kang, C. and Vermerris, W. (2009) A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum (*Sorghum bicolor* (L.) Moench) identifies *SbCAD2* as the *Brown midrib6* gene. *Genetics*, **181**, 783–795.

Sattler, S.E., Saathoff, A.J., Haas, E.J., Palmer, N.A., Funnell-Harris, D.L., Sarath, G. and Pedersen, J.F. (2009) A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum *brown midrib6* phenotype. *Plant Physiol.* **150**, 584–595.

Sattler, S.E., Funnell-Harris, D.L. and Pedersen, J.F. (2010) *Brown midrib* mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. *Plant Sci.* **178**, 229–238.

Schmittgen, T.D. (2006) Quantitative gene expression by real-time PCR: a complete protocol. In *Real-Time PCR* (Dorak, M.T., ed). New York: Taylor and Francis, pp. 127–137.

Schulz, A.R. (1994) *Enzyme Kinetics: From Diastase to Multi-Enzyme Systems*. Cambridge: Cambridge University Press.

Souza, C.d.A., Barbazuk, B., Ralph, S.G., Bohlmann, J., Hamberger, B., Douglas, C.J. and de A. Souza, C. (2008) Genome-wide analysis of a land plant-specific acyl:coenzymeA synthetase (ACS) gene family in Arabidopsis, poplar, rice and *Physcomitrella*. *New Phytol.* **179**, 987–1003.

Stracke, R., Werber, M. and Weisshaar, B. (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**, 447–456.

Takano, S., Niihama, M., Smith, H.M.S., Tasaka, M. and Aida, M. (2010) *gorgon*, a novel missense mutation in the *SHOOT MERISTEMLESS* gene, impairs shoot meristem homeostasis in Arabidopsis. *Plant Cell Physiol.* **51**, 621–634.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Vanholme, R., Morreel, K., Ralph, J. and Boerjan, W. (2008) Lignin engineering. *Curr. Opin. Plant Biol.* **11**, 278–285.

Vermerris, W., Saballos, A., Ejeta, G., Mosier, N.S., Ladisch, M.R. and Carpita, N.C. (2007) Molecular breeding to enhance ethanol production from corn and sorghum stover. *Crop Sci.* **47**, S142–S153.

Vermerris, W., Sherman, D.M. and McIntyre, L.M. (2010) Phenotypic plasticity in cell walls of maize *brown midrib* mutants is limited by lignin composition. *J. Exp. Bot.* **61**, 2479–2490.

Voelker, S.L., Lachenbruch, B., Meinzer, F.C. *et al.* (2010) Antisense down-regulation of *4CL* expression alters lignification, tree growth, and saccharification potential of field-grown poplar. *Plant Physiol.* **154**, 874–886.

Xin, Z., Wang, M.L., Barkley, N.A., Burow, G., Franks, C., Pederson, G. and Burke, J. (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC Plant Biol.* **8**, 103.

Xu, Z., Zhang, D., Hu, J. *et al.* (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics*, **10**(Suppl. 11), S3.

Xu, B., Escamilla-Treviño, L.L., Sathitsuksanoh, N., Shen, Z., Shen, H., Zhang, Y.-H., Dixon, R.A. and 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.* **192**, 611–625.

Yokoyama, R. and Nishitani, K. (2004) Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and Arabidopsis. *Plant Cell Physiol.* **45**, 1111–1121.