Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2940141 C 2022/07/12

(11)(21) **2 940 141**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2014/03/11
(87) Date publication PCT/PCT Publication Date: 2014/09/25
(45) Date de délivrance/Issue Date: 2022/07/12
(85) Entrée phase nationale/National Entry: 2016/08/18
(86) N° demande PCT/PCT Application No.: US 2014/023443
(87) N° publication PCT/PCT Publication No.: 2014/150504
(30) Priorité/Priority: 2013/03/15 (US61/792,864)

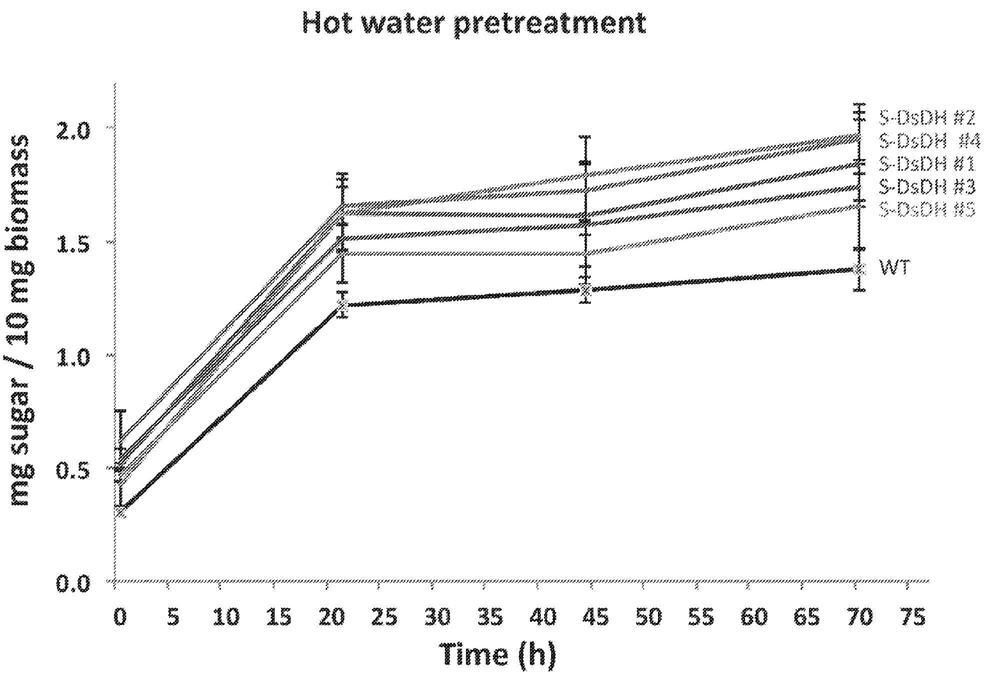
(51) Cl.Int./Int.Cl. *C12N 5/10* (2006.01), *A01H 5/00* (2018.01), *C12N 15/52* (2006.01), *C12N 15/54* (2006.01), *C12N 15/60* (2006.01), *C12N 15/82* (2006.01), *C12N 9/12* (2006.01), *C12N 9/88* (2006.01), *C12P 19/00* (2006.01), *C13K 1/00* (2006.01)

(72) Inventeurs/Inventors:
LOQUE, DOMINIQUE, US;
EUDES, AYMERICK, US

(73) Propriétaire/Owner:
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, US

(74) Agent: SMART & BIGGAR LLP

(54) Titre : REDUCTION SPECIFIQUE AU TISSU DE LA LIGNINE
(54) Title: TISSUE SPECIFIC REDUCTION OF LIGNIN



(57) Abrégé/Abstract:
The present invention provides an expression cassette comprising a polynucleotide that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway in the plant, which is operably linked to a heterologous promoter. Also provided are methods of engineering a plant having reduced lignin content, as well as plant cells, plant parts, and plant tissues from such engineered plants.

CA 02940141 2016-08-18

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 September 2014 (25.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/150504 A3

(51) International Patent Classification:
A01H 5/00 (2006.01) *C12N 15/82* (2006.01)

(21) International Application Number:
PCT/US2014/023443

(22) International Filing Date:
11 March 2014 (11.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/792,864 15 March 2013 (15.03.2013) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors: **LOQUE, Dominique**; 974 Neilson Street, Albany, California 94706 (US). **EUDES, Aymerick**; 6363 Christie Avenue, Emeryville, California 94608 (US).

(74) Agents: **FLANAGAN, Lisa Dornbach** et al.; Kilpatrick Townsend & Stockton LLP, Eighth Floor, Two Embarcadero Center, San Francisco, California 94111 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

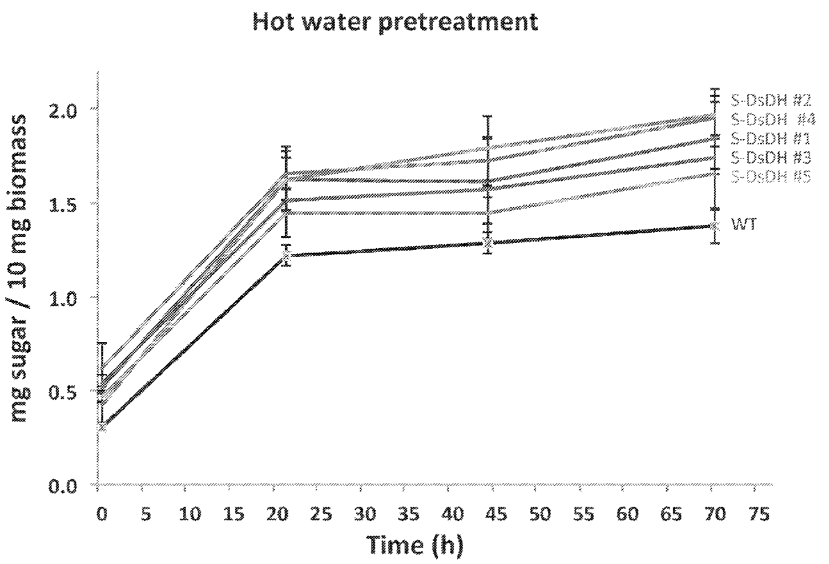
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(88) Date of publication of the international search report:
6 November 2014

(54) Title: TISSUE SPECIFIC REDUCTION OF LIGNIN

FIG. 15



(57) Abstract: The present invention provides an expression cassette comprising a polynucleotide that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway in the plant, which is operably linked to a heterologous promoter. Also provided are methods of engineering a plant having reduced lignin content, as well as plant cells, plant parts, and plant tissues from such engineered plants.

WO 2014/150504 A3

TISSUE SPECIFIC REDUCTION OF LIGNIN

[0001] [deleted]

[0002] [deleted]

BACKGROUND OF THE INVENTION

[0003] Plant lignocellulosic biomass is used as a renewable feedstock for biofuel production and is a promising alternative to fossil fuel consumption. However, a major bottleneck in biofuel production is the quality of available feedstocks. Available feedstocks have a high resistance (recalcitrance) to being reduced into simple sugars that can in turn be converted into fuel. Therefore, improving the composition and/or digestibility of the raw biomass will have an important beneficial impact on lignocellulosic biofuels production.

[0004] Lignocellulosic biomass is mainly composed of secondary cell walls, which comprise polysaccharide polymers embedded in lignin. The embedding of the polysaccharide polymers in lignin reduces their extractability and accessibility to hydrolytic enzymes, resulting in cell wall recalcitrance to enzymatic hydrolysis. Lignin content and saccharification efficiency of plant cell wall usually are highly negatively correlated. *See, e.g.,* Chen and Dixon, *Nat. Biotechnol.* 25:759-761 (2007); Jorgensen *et al., Biofuel Bioprod. Bior.* 1:119-134 (2007); and Vinzant *et al., Appl. Biochem. Biotechnol.* 62:99-104 (1997). However, most attempts at reducing lignin content during plant development have resulted in severe biomass yield reduction (Franke *et al., Plant J.* 30:33-45 (2002); Shadle *et al., Phytochemistry* 68:1521-1529 (2007); and Voelker *et al., Plant Physiol.* 154:874-886 (2010)) and therefore, there are few crops having significant lignin reduction. Although silencing

Date Reçue/Date Received 2020-05-29

strategies have been used to reduce the amount of lignin in plants, there remains a need for methods of reducing lignin in specific cell and tissue types that reduce cell wall recalcitrance, thus improving the extractability and hydrolysis of fermentable sugars from plant biomass.

BRIEF SUMMARY OF THE INVENTION

[0005] In one aspect, the present invention provides methods of engineering a plant having reduced lignin content. In some embodiments, the method comprises:

introducing into the plant an expression cassette comprising a polynucleotide that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway (*e.g.*, a p-coumaryl alcohol, sinapyl alcohol, and/or coniferyl alcohol biosynthesis pathway) in the plant, and wherein the polynucleotide is operably linked to a heterologous promoter; and

culturing the plant under conditions in which the protein that diverts the monolignol precursor from the lignin biosynthesis pathway is expressed.

[0006] In some embodiments, the protein reduces the amount of cytosolic and/or plastidial shikimate that is available for the lignin biosynthesis pathway. In some embodiments, the protein is shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), or dehydroshikimate dehydratase (QsuB). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

[0007] In some embodiments, the protein reduces the amount of cytosolic and/or plastidial phenylalanine that is available for the lignin biosynthesis pathway. In some embodiments, wherein the protein is phenylacetaldehyde synthase (PAAS) or phenylalanine aminomutase (PAM). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:10 or SEQ ID NO:29.

[0008] In some embodiments, the protein reduces the amount of cinnamate and/or coumarate that is available for the lignin biosynthesis pathway. In some embodiments, the protein is p-coumarate/cinnamate carboxylmethltransferase (CCMT1) or phenylacrylic acid decarboxylase (PDC). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:12 or SEQ ID NO:30.

[0009] In some embodiments, the protein reduces the amount of coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the lignin biosynthesis pathway. In some embodiments, the protein is 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:14, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO;35, or SEQ ID NO:36.

[0010] In some embodiments, the protein activates or potentiates a metabolic pathway that competes with the lignin biosynthesis pathway for the use of monolignol precursors. In some embodiments, the metabolic pathway is a stilbene biosynthesis pathway, a flavonoid biosynthesis pathway, a curcuminoid biosynthesis pathway, or a bensalacetone biosynthesis pathway. In some embodiments, the protein is a transcription factor that activates or potentiates the flavonoid biosynthesis pathway. In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, or SEQ ID NO:45.

[0011] In some embodiments, the promoter is a tissue-specific promoter. In some embodiments, the promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter. In some embodiments, the promoter is an IRX5 promoter. In some embodiments, the promoter is from a gene that is co-expressed in the lignin biosynthesis pathway (phenylpropanoid pathway), *e.g.*, a promoter from a gene expressed in the pathway shown in Figure 1. In some embodiments, the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter.

[0012] In some embodiments, the protein that diverts a monolignol precursor from a lignin biosynthesis pathway is targeted to a plastid in the plant. In some embodiments, the polynucleotide comprises a plastid targeting signal that is substantially identical to the polynucleotide sequence of SEQ ID NO:15.

[0013] In some embodiments, the protein diverts a monolignol precursor from a sinapyl alcohol and/or coniferyl alcohol biosynthesis pathway. In some embodiments, the plant has reduced content of guaiacyl (G) and syringyl (S) lignin units.

[0014] In some embodiments, the plant (or plant part, or seed, flower, leaf, or fruit from the plant) is selected from the group consisting of Arabidopsis, poplar, eucalyptus, rice, corn, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, and Brachypodium.

[0015] In another aspect, the present invention provides a plant cell comprising a polynucleotide that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway in the plant, wherein the polynucleotide is operably linked to a heterologous promoter.

[0016] In some embodiments, the plant cell comprises a polynucleotide that encodes a protein that reduces the amount of cytosolic and/or plastidial shikimate that is available for the lignin biosynthesis pathway. In some embodiments, the protein is shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), or dehydroshikimate dehydratase (QsuB). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

[0017] In some embodiments, the plant cell comprises a polynucleotide that encodes a protein that reduces the amount of cytosolic and/or plastidial phenylalanine that is available for the lignin biosynthesis pathway. In some embodiments, wherein the protein is phenylacetaldehyde synthase (PAAS) or phenylalanine aminomutase (PAM). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:10 or SEQ ID NO:29.

[0018] In some embodiments, the plant cell comprises a polynucleotide that encodes a protein that reduces the amount of cinnamate and/or coumarate that is available for the lignin biosynthesis pathway. In some embodiments, the protein is p-coumarate/cinnamate carboxylmethltransferase (CCMT1) or phenylacrylic decarboxylase (PDC). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:12 or SEQ ID NO:30.

[0019] In some embodiments, the plant cell comprises a polynucleotide that encodes a protein that reduces the amount of coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the lignin biosynthesis pathway. In some embodiments, the protein is 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS). In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:14, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO;35, or SEQ ID NO:36.

[0020] In some embodiments, the plant cell comprises a polynucleotide that encodes a protein activates or potentiates a metabolic pathway that competes with the lignin biosynthesis pathway for the use of monolignol precursors. In some embodiments, the metabolic pathway is a stilbene biosynthesis pathway, a flavonoid biosynthesis pathway, a curcuminoid biosynthesis pathway, or a bensalacetone biosynthesis pathway. In some embodiments, the protein is a transcription factor that activates or potentiates the flavonoid biosynthesis pathway. In some embodiments, the protein is substantially identical to an amino acid sequence of SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, or SEQ ID NO:45.

[0021] In some embodiments, the plant cell comprises a tissue-specific promoter. In some embodiments, the promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter. In some embodiments, the promoter is an IRX5 promoter. In some embodiments, the plant cell comprises a promoter from a gene that is co-expressed in the lignin biosynthesis pathway (phenylpropanoid pathway), *e.g.*, a promoter from a gene expressed in the pathway shown in Figure 1. In some embodiments, the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter.

[0022] In some embodiments, the plant cell comprises a polynucleotide encoding a protein that diverts a monolignol precursor from a lignin biosynthesis pathway that is targeted to a plastid in the plant. In some embodiments, the polynucleotide comprises a plastid targeting signal that is substantially identical to the polynucleotide sequence of SEQ ID NO:15.

[0023] In another aspect, the present invention provides plants comprising a plant cell as described herein. In some embodiments, the plant has reduced lignin content that is substantially localized to secondary cell wall tissue or fiber cells of the plant.

[0024] In yet another aspect, the present invention provides methods of engineering a plant having reduced lignin content by expressing or overexpressing a competitive inhibitor of a lignin biosynthesis pathway enzyme. In some embodiments, the method comprises:

introducing into the plant an expression cassette comprising a polynucleotide that encodes a protein that produces a competitive inhibitor of hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) in the plant, wherein the polynucleotide is operably linked to a heterologous promoter; and

culturing the plant under conditions in which the protein that produces a competitive inhibitor of HCT is expressed.

[0025] In some embodiments, the protein produces one or more of the competitive inhibitors protocatechuate, gentisate, catechol, 2,3-dihydroxybenzoate, 3,6-dihydroxybenzoate, or 3-hydroxy-2-aminobenzoate. In some embodiments, the protein produces the competitive inhibitor of HCT protocatechuate. In some embodiments, the protein is dehydroshikimate dehydratase (QsuB), dehydroshikimate dehydratase (DsDH), isochorismate synthase (ICS), salicylic acid 3-hydroxylase (S3H), salicylate hydroxylase (nahG), or salicylate 5-hydroxylase (nagGH).

[0026] In some embodiments, the polynucleotide that encodes a protein that produces a competitive inhibitor of HCT is operably linked to a tissue-specific promoter. In some embodiments, the promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter. In some embodiments, the promoter is an IRX5 promoter. In some embodiments, the promoter is from a gene that is expressed in the lignin biosynthesis pathway (phenylpropanoid pathway), *e.g.*, a promoter from a gene expressed in the pathway shown in Figure 1. In some embodiments, the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter.

[0027] In still another aspect, the present invention provides a plant, plant part, or seed, flower, leaf, or fruit from the plant, or a plant cell comprising a polynucleotide that encodes a protein that produces a competitive inhibitor of HCT in the plant, wherein the polynucleotide is operably linked to a heterologous promoter.

[0028] In still another aspect, the present invention provides biomass comprising plant tissue from a plant or part of a plant as described herein.

[0029] In yet another aspect, the present invention provides methods of obtaining an increased amount of soluble sugars from a plant in a saccharification reaction. In some embodiments, the method comprises subjecting a plant as described herein to a saccharification reaction, thereby increasing the amount of soluble sugars that can be obtained from the plant as compared to a wild-type plant.

[0030] In still another aspect, the present invention provides methods of increasing the digestibility of the biomass for ruminants. In some embodiments, the method comprises introducing an expression cassette as described herein into a plant; culturing the plant under conditions in which the protein that diverts the monolignol precursor from the lignin biosynthesis pathway, or the protein that produces a competitive inhibitor of HCT, is expressed; and obtaining biomass from the plant, thereby increasing the digestibility of the biomass for ruminants.

[0030A] Aspects of the disclosure relate to a method of engineering a plant having reduced lignin content, the method comprising: introducing into the plant an expression cassette comprising a polynucleotide that encodes: a bacterial dehydroshikimate dehydratase or a *Podospora anserina* dehydroshikimate dehydratase (DsDH); a bacterial shikimate kinase; a pentafunctional AROM polypeptide (ARO1) a phenylacetaldehyde synthase (PAAS) or phenylalanine aminomutase (PAM); a p-coumarate/cinnamate carboxylmethltransferase (CCMT1) or phenylacrylic acid decarboxylase (PDC); a 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS); or a transcription factor that activates or potentiates the flavonoid biosynthesis pathway; wherein the polynucleotide is operably linked to a heterologous promoter and wherein the heterologous promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter; and culturing the plant under conditions in which the protein that diverts the monolignol precursor from the lignin biosynthesis pathway is expressed.

[0030B] Aspects of the disclosure relate to a plant cell comprising an expression cassette comprising a polynucleotide that encodes: a bacterial dehydroshikimate dehydratase or a *Podospora anserina* dehydroshikimate dehydratase (DsDH); a bacterial shikimate kinase; a pentafunctional AROM polypeptide (ARO1)-a phenylacetaldehyde synthase (PAAS) or phenylalanine aminomutase (PAM); a p-coumarate/cinnamate carboxylmethltransferase (CCMT1) or phenylacrylic acid decarboxylase (PDC); a 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS); or a transcription factor that activates or potentiates the flavonoid biosynthesis pathway; wherein the polynucleotide is operably linked to a heterologous promoter and wherein the heterologous promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter.

[0030C] Aspects of the disclosure relate to a plant cell comprising a polynucleotide that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway in the plant, wherein the polynucleotide is operably linked to a heterologous promoter, wherein the plant cell is from secondary cell wall or fiber that diverts a monolignol precursor from the lignin pathway is shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), dehydroshikimate dehydratase (QsuB), phenylacetaldehyde synthase (PAAS), phenylalanine aminomutase (PAM), p-coumarate/cinnamate carboxylmethltransferase (CCMT1), ferulic acid decarboxylase (FDC1), phenylacrylic decarboxylase (PDC), 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS); and the promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter.

Date Reçue/Date Received 2021-05-05

[0030D] Various embodiments of the claimed invention relate to a method of engineering a plant having reduced lignin content, the method comprising: introducing into the plant an expression cassette comprising a polynucleotide that encodes an enzyme that diverts a monolignol precursor from the lignin biosynthesis pathway, wherein the polynucleotide is operably linked to a heterologous secondary cell wall-specific promoter or a heterologous fiber cell-specific promoter, and wherein the enzyme is: a bacterial dehydroshikimate dehydratase, a *Podospora anserina* dehydroshikimate dehydratase (DsDH), a bacterial shikimate kinase, a pentafunctional AROM polypeptide (ARO1), a phenylacetaldehyde synthase (PAAS), a phenylalanine aminomutase (PAM), a *p*-coumarate/cinnamate carboxylmethlytransferase (CCMT1), a phenylacrylic acid decarboxylase (PDC), a 2-oxoglutarate-dependent dioxygenase (C2'H), a chalcone synthase (CHS), a stilbene synthase (SPS), a cucuminoid synthase (CUS), or a benzalacetone synthase (BAS); and culturing the plant under conditions in which the enzyme that diverts the monolignol precursor from the lignin biosynthesis pathway is expressed, thereby reducing the plant lignin content.

[0030E] Various embodiments of the claimed invention also relate to a plant cell comprising an expression cassette comprising a polynucleotide that encodes an enzyme that diverts a monolignol precursor from the lignin biosynthesis pathway, wherein the polynucleotide is operably linked to a heterologous secondary cell wall-specific promoter or a heterologous fiber cell-specific promoter, and wherein the enzyme is a bacterial dehydroshikimate dehydratase, a *Podospora anserina* dehydroshikimate dehydratase (DsDH), a bacterial shikimate kinase, a pentafunctional AROM polypeptide (ARO1), a phenylacetaldehyde synthase (PAAS), a phenylalanine aminomutase (PAM), a *p*-coumarate/cinnamate carboxylmethyltransferase (CCMT1), a phenylacrylic acid decarboxylase (PDC), a 2-oxoglutarate-dependent dioxygenase (C2'H), a chalcone synthase (CHS), a stilbene synthase (SPS), a cucuminoid synthase (CUS), or a benzalacetone synthase (BAS).

[0030F] Various embodiments of the claimed invention also relate to a plant cell comprising a polynucleotide that encodes an enzyme that diverts a monolignol precursor from a lignin biosynthesis pathway in a plant, wherein: the polynucleotide is operably linked to a heterologous promoter; the plant cell is a secondary cell wall or fiber cell; the enzyme is shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), dehydroshikimate dehydratase (QsuB), phenylacetaldehyde synthase (PAAS), phenylalanine aminomutase (PAM), *p*-coumarate/cinnamate carboxylmethltransferase (CCMT1), ferulic acid decarboxylase (FDC1), phenylacrylic decarboxylase (PDC), 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone synthase (BAS); and the heterologous promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter.

Date Reçue/Date Received 2021-05-05

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] **Figure 1.** Representation of the lignin biosynthesis pathway. Modified lignin biosynthesis pathway from Fraser and Chapple (2011). Enzyme descriptions: PAL: phenylalanine ammonia-lyase; C4H: cinnamate-4-hydroxylase; 4CL: 4-hydroxycinnamate CoA-ligase; HCT: hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase; C3'H: 4-hydroxycinnamate 3-hydroxylase; CCoAOMT: caffeoyl-CoA O-methyltransferase; CCR: hydroxycinnamoyl-CoA NADPH oxidoreductase; COMT: caffeate O-methyltransferase; CAD: hydroxycinnamyl alcohol dehydrogenase; F5H: ferulate 5-hydroxylase. Name of the lignin precursors: 1, phenylalanine; 2, cinnamate; 3, p-coumarate; 4, p-coumaroyl-CoA; 5, *p*-coumaroyl-shikimate/quinate (R = shikimate/quinate); 6, caffeoyl-shikimate/quinate; 7, caffeoyl-CoA; 8, feruloyl-CoA; 9, p-coumaraldehyde; 10, coniferaldehyde; 11, 5-hydroxy- coniferaldehyde; 12, sinapaldehyde; 13, p-coumaryl alcohol; 14, coniferyl alcohol; 15, sinapyl alcohol.

[0032] **Figure 2.** Lignin reduction via depletion of shikimate (HCT co-substrate). Strategies for reducing or depleting the amount of shikimate that is available for the lignin biosynthesis pathway are shown. (1) The amount of cytosolic shikimate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a shikimate kinase such as *M. tuberculosis* shikimate kinase ("MtAroK"). (2) The amount of plastidial shikimate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a pentafunctional arom protein such as *S. cerevisiae* pentafunctional arom protein ("ScAro1"). Plastidial expression of the protein can be accomplished via a plastid targeting signal, *e.g.*, as described herein.

[0033] **Figure 3.** Lignin reduction via depletion of shikimate and production of new stoppers. Strategies for reducing or depleting the amount of shikimate that is available for the lignin biosynthesis pathway are shown. For example, the amount of plastidial shikimate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a dehydroshikimate dehydratase such as *C. glutamicum* dehydroshikimate dehydratase ("CgQsuB") or *P. anserina* dehydroshikimate dehydratase ("PaDsDH"). Plastidial expression of the protein can be accomplished via a plastid targeting signal, *e.g.*, as described herein.

[0034] **Figure 4.** Lignin reduction via depletion of phenylalanine (PAL substrate). Strategies for reducing or depleting the amount of phenylalanine that is available for the lignin biosynthesis pathway are shown. For example, the amount of (1) cytosolic and/or (2) plastidial phenylalanine that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a phenylacetaldehyde such as *P.hybrida* phenylacetaldehyde synthase ("PhPAAS"). Plastidial expression of the protein can be accomplished via a plastid targeting signal, *e.g.*, as described herein.

[0035] **Figure 5.** Lignin reduction via depletion of phenylalanine (PAL substrate). Strategies for reducing or depleting the amount of phenylalanine that is available for the lignin biosynthesis pathway are shown. For example, the amount of (1) cytosolic and/or (2) plastidial phenylalanine that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a phenylalanine aminomutase such as *T. canadensis* phenylalanine aminomutase ("TcPAM"). Plastidial expression of the protein can be accomplished via a plastid targeting signal, *e.g.*, as described herein.

[0036] **Figure 6.** Lignin reduction via depletion of cinnamate (C4H substrate) and coumarate (4CL substrate). Strategies for reducing or depleting the amount of cinnamate and/or p-coumarate that is available for the lignin biosynthesis pathway are shown. For example, the amount of cytosolic cinnamate and/or p-coumarate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a cinnamate/p-coumarate carboxyl methyltransferase such as *O. basilicum* cinnamate/p-coumarate carboxyl methyltransferase ("ObCCMT1").

[0037] **Figure 7.** Lignin reduction via depletion of cinnamate (C4H substrate) and coumarate (4CL substrate). Strategies for reducing or depleting the amount of cinnamate and/or p-coumarate that is available for the lignin biosynthesis pathway are shown. For example, the amount of cytosolic cinnamate and/or p-coumarate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a phenylacrylic decarboxylase (PDC or PAD).

[0038] **Figure 8.** Lignin reduction via depletion of coumaroyl-CoA (HCT substrate). Strategies for reducing or depleting the amount of coumaroyl-CoA that is available for the lignin biosynthesis pathway are shown. For example, the amount of cytosolic coumaroyl-CoA that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a 2-oxoglutarate-dependent dioxygenase such as *R. graveolens* C2'H (2-oxoglutarate-dependent dioxygenase) ("RbC2'H").

[0039] **Figure 9.** Lignin reduction via depletion of coumaroyl-CoA (HCT substrate). Strategies for reducing or depleting the amount of coumaroyl-CoA that is available for the lignin biosynthesis pathway are shown. For example, the amount of cytosolic coumaroyl-CoA that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS).

[0040] **Figure 10.** Lignin reduction via depletion of feruloyl-CoA (CCR substrate). Strategies for reducing or depleting the amount of feruloyl-CoA that is available for the lignin biosynthesis pathway are shown. For example, the amount of cytosolic feruloyl-CoA that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a 2-oxoglutarate-dependent dioxygenase such as *R. graveolens* C2'H (2-oxoglutarate-dependent dioxygenase) ("RbC2'H").

[0041] **Figure 11.** Lignin reduction via depletion of caffeoyl-CoA feruloyl-CoA (CCR substrate). Strategies for reducing or depleting the amount of caffeoyl-CoA and/or feruloyl-CoA that is available for the lignin biosynthesis pathway are shown. For example, the amount of cytosolic caffeoyl-CoA and/or feruloyl-CoA that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a chalcone synthase (CHS), synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS).

[0042] **Figure 12.** Growth phenotype analysis of S-QsuB lines. Picture of 3 weeks-old plants at rosette stage. No phenotypic differences could be observed between S-QsuB lines and WT plants at the rosette stage.

[0043] **Figure 13.** Total reducing-sugars released from stem biomass of S-QsuB lines and WT plants after 72h incubation with a cellulolytic enzyme cocktail. Total reducing-sugars released from biomass after hot-water pretreatment (1h at 120C) and incubation with a cellulolytic enzyme cocktail (Novozymes Cellic® CTec2) at a loading of 0.88% (g enzyme / g biomass) were measured using the 3,5-Dinitrosalicylic acid assay as described in Eudes *et al.* 2012 (*Plant Biotech Journal* 10(5):609-620).

[0044] **Figure 14.** Time course for total reducing-sugars released from stem biomass of S-QsuB lines and WT plants after incubation with different loadings of a cellulolytic enzyme cocktail. Time course for total reducing-sugars released from biomass after hot-water pretreatment (1h at 120C) and incubation with different loadings (0.88%, 0.176% or 0.088%; g of enzyme / g of biomass) of a cellulolytic enzyme cocktail (Novozymes Cellic® CTec2) . Measurements were performed as described in (Eudes *et al.* 2012 Plant Biotech Journal 10(5):609-620).

[0045] **Figure 15.** Total reducing-sugars released from stem biomass of S-DsDH lines after 72h incubation with a cellulolytic enzyme cocktail. Time course for total reducing-sugar released from biomass after hot-water pretreatment (1h at 120C) and incubation with a cellulolytic enzyme cocktail (Novozymes Cellic® CTec2) at a loading of 0.88% (g enzyme / g biomass). Measurements were performed as described in (Eudes *et al.* 2012 Plant Biotech Journal 10(5):609-620).

[0046] Figure 16. QsuB expression in Arabidopsis stems. Detection by Western blot of QsuB tagged with the AttB2 peptide (approximate size 70 kDa) using the "universal antibody" and stem proteins from nine independent 6-wk-old *pC4H::schl::qsuB* (*C4H::qsuB*) T2 transformants. A stem protein extract from wild type was used as a negative control (WT) and a Ponceau staining of Rubisco large subunit (rbcL) is shown as a loading control.

[0047] Figure 17. Partial short-range ^{13}C–^{1}H (HSQC) spectra (aromatic region) of cell-wall material from mature senesced stems of wild-type (WT), *pC4H::schl::qsuB-1* (*C4H::qsuB-1*) and *pC4H::schl::qsuB-9* (*C4H::qsuB-9*) plants. Lignin monomer ratios are provided on the figures.

[0048] Figure 18. Polydispersity of cellulolytic enzyme lignins from wild-type and *C4H::qsuB* lines. Cellulolytic enzyme lignins were purified from mature senesced stems of wild-type (WT, black line), *pC4H::schl::qsuB-1* (*C4H::qsuB-1*, red line) and *pC4H::schl::qsuB-9* (*C4H::qsuB-9*, purple line) plants and analyzed for polydispersity by size-exclusion chromatography (SEC). SEC chromatograms were obtained using UV-F fluorescence (Ex_{250}/Em_{450}). m, molecular weight.

[0049] Figure 19. Saccharification of biomass from mature senesced stems of wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) lines. (A) Amounts of sugars released from biomass after various pretreatments and 72-h enzymatic digestion with cellulase (1% w/w). Values are means ± SE of four biological replicates ($n = 4$). Asterisks indicate significant differences from the wild type using the unpaired Student's t-test (*$P < 0.05$; **$P < 0.005$). (B) Amounts of sugars released from biomass after hot water pretreatment and 72-h enzymatic digestion using two different cellulase loadings (1% or 0.2% w/w). Values are means ± SE of four biological replicates ($n = 4$). Asterisks indicate significant differences from the wild type at 1% cellulase loading using the unpaired Student's t-test (*$P < 0.05$; **$P < 0.005$).

[0050] Figure 20. The lignin biosynthetic pathway. Abbreviations: DAHPS, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase; DHQS, 3-dehydroquinate synthase; DHQD/SD, 3-dehydroquinate dehydratase; SK, shikimate kinase; ESPS, 3-phosphoshikimate 1-carboxyvinyltransferase; CS, chorismate synthase; CM, chorismate mutase; PAT, prephenate aminotransferase; ADT, arogenate dehydratase; PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; CSE, caffeoyl shikimate esterase; 4CL, 4-coumarate CoA ligase; CAD, cinnamyl alcohol dehydrogenase; F5H, ferulate 5-hydroxylase; C3H, coumarate 3-hydroxylase; COMT, caffeic acid 3-O-methyltransferase; CCR, cinnamoyl-CoA reductase; HCT, hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase; CCoAOMT, caffeoyl/CoA-3-*O*-methyltransferase; qsuB, 3-dehydroshikimate dehydratase from *Corynebacterium glutamicum*.

[0051] **Figure 21.** Subcellular localization of SCHL-QsuB. The left panel displays the transient expression of SCHL-QsuB-YFP fusion protein expressed under the control of the 35S promoter in epidermal cells of *N. benthamiana* and imaged by confocal laser scanning microscopy. The central panel displays fluorescing chloroplasts and the right panel shows the merged images (colocalizations are visible as yellow dots). Scale bars = 20 μm.

[0052] **Figure 22.** Summary of the fold changes observed for the methanol-soluble metabolites extracted from plants expressing QsuB.

[0053] **Figure 23.** Partial short-range ^{13}C–^{1}H (HSQC) spectra (aliphatic region) of cell wall material from mature senesced stems of wild-type (WT), *pC4H::schl::qsuB-1* (*C4H::qsuB-1*) and *pC4H::schl::qsuB-9* (*C4H::qsuB-9*) plants.

[0054] **Figure 24.** Lignin staining by phloroglucinol-HCl of stem sections from 5-wk-old wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) plants.

[0055] **Figure 25.** LC-MS chromatograms from AtHCT *in-vivo* activity assays. LC-MS chromatograms of coumarate conjugates produced by AtHCT after feeding a recombinant yeast strain co-expressing At4CL5 and AtHCT with *p*-coumarate and (A) shikimate, (B) 3,6-dihydroxybenzoate, (C) 3-hydroxy-2-amino benzoate, (D) 2,3-dihydroxybenzoate, (E) catechol, or (F) protocatechuate are presented. Structures of coumarate-dihydroxybenzoate esters are arbitrary shown with an ester linkage at the 3-hydroxy position of the dihydroxybenzoate ring. The structure of coumaroyl-3-hydroxyanthranilate (*C*) is represented as determined in Moglia et al. (34).

[0056] **Figure 26.** LC-MS chromatogram of *p*-coumaraldehyde detected in methanol-soluble extracts of stems from lines expressing QsuB.

[0057] **Figure 27.** Competitive inhibitor pathways.

[0058] **Figure 28.** Characteristics and relative molar abundances (%) of the compounds released after pyro-GC/MS of extractive-free senesced mature stems from wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) plants. Values in brackets are the SE from duplicate analyses. nd, not detected.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

[0059] As used herein, the term "lignin biosynthesis pathway" refers to an enzymatic pathway (the phenylpropanoid pathway) in plants in which the lignin monomers (p-coumaryl (4-hydroxycinnamyl) alcohol, coniferyl (3-methoxy 4-hydroxycinnamyl) alcohol, and sinapyl (3,5-dimethoxy 4-hydroxycinnamyl) alcohol) are synthesized from phenylalanine. The lignin biosynthesis pathway and enzymatic components of the pathway are depicted, for example, in Figure 1.

[0060] As used herein, the term "monolignol precursor" refers to a substrate of the lignin biosynthesis pathway that is directly or indirectly synthesized into a lignin monomer. In some embodiments, a monolignol precursor is a substrate of the lignin biosynthesis pathway that is identified in any of Figures 1-11.

[0061] As used herein, the term "protein that diverts a monolignol precursor from a lignin biosynthesis pathway" refers to a protein that activates, promotes, potentiates, or enhances expression of an enzymatic reaction or metabolic pathway that decreases the amount of monolignol precursor that is available for the synthesis of a lignin monomer. The term includes polymorphic variants, alleles, mutants, and interspecies homologs to the specific proteins (*e.g.*, enzymes) described herein. A nucleic acid that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway (or a nucleic acid that encodes a protein that diverts a monolignol precursor from a p-coumaryl alcohol, sinapyl alcohol, and/or coniferyl alcohol pathway) refers to a gene, pre-mRNA, mRNA, and the like, including nucleic acids encoding polymorphic variants, alleles, mutants, and interspecies homologs of the particular proteins (*e.g.*, enzymes) described herein. In some embodiments, a nucleic acid that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway (1) has a nucleic acid sequence that has greater than about 50% nucleotide sequence identity, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or higher nucleotide sequence identity, preferably over a region of at least about 10, 15, 20, 25, 50, 100, 200, 500 or more nucleotides or over the length of the entire polynucleotide, to a nucleic acid sequence of any of SEQ ID NOs:1, 3, 5, 7, 9, 11, or 13; or (2) encodes a polypeptide having an amino acid sequence that has greater than about 50% amino acid sequence identity, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity, preferably over a region of at least about 25, 50, 100, 200 or more amino acids or over the length of the entire polypeptide, to a polypeptide encoded by a nucleic acid sequence of any of SEQ ID NOs:1, 3, 5, 7, 9, 11, or 13, or to an amino acid sequence of any of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 42, 43, 44, or 45. In some embodiments, a protein that diverts a monolignol precursor from a lignin biosynthesis pathway has an amino acid sequence having greater than about 50% amino acid sequence identity, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity, preferably over a region of at least about 25, 50, 100, 200 or more amino acids or over the length of the entire polypeptide, to an amino acid sequence of any of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 42, 43, 44, or 45.

[0062] The term "protein that produces a competitive inhibitor of HCT" refers to a protein that directly or indirectly produces a molecule that can compete with *p*-coumaroyl-CoA and/or shikimate as a substrate for hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT), thereby acting as a competitive inhibitor of HCT. Non-limiting examples of molecules (e.g., metabolites) that can act as competitive inhibitors of HCT are shown in Figure 27. In some embodiments, the competitive inhibitor of HCT is protocatechuate, catechol, 3,6-dihydroxybenzoate, 3-hydroxy-2-aminobenzoate, or 2,3-dihydroxybenzoate. Thus, in some embodiments, the protein that produces a competitive inhibitor of HCT is a protein (*e.g.,* an enzyme) that directly or indirectly produces protocatechuate, catechol, 3,6-dihydroxybenzoate, 3-hydroxy-2-aminobenzoate, or 2,3-dihydroxybenzoate, including but not limited to the enzymes dehydroshikimate dehydratase (QsuB), dehydroshikimate dehydratase (DsDH), isochorismate synthase (ICS), salicylic acid 3-hydroxylase (S3H), salicylate hydroxylase (nahG), and salicylate 5-hydroxylase (nagGH). In some embodiments, an *in vivo* enzymatic assay, for example as described in the Examples section below, can be used to determine whether a molecule can compete with *p*-coumaroyl-CoA and/or shikimate as a substrate for HCT.

[0063] The terms "polynucleotide" and "nucleic acid" are used interchangeably and refer to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs may be used that may have alternate backbones, comprising, *e.g.*, phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press); positive backbones; non-ionic backbones, and non-ribose backbones. Thus, nucleic acids or polynucleotides may also include modified nucleotides that permit correct read-through by a polymerase. "Polynucleotide sequence" or "nucleic acid sequence" includes both the sense and antisense strands of a nucleic acid as either individual single strands or in a duplex. As will be appreciated by those in the art, the depiction of a single strand also defines the sequence of the complementary strand; thus the sequences described herein also provide the complement of the sequence. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc.

[0064] The term "substantially identical," used in the context of two nucleic acids or polypeptides, refers to a sequence that has at least 50% sequence identity with a reference sequence. Percent identity can be any integer from 50% to 100%. Some embodiments include at least: 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below. For example, a first polynucleotide is substantially identical to a second polynucleotide sequence if the first polynucleotide sequence is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the second polynucleotide sequence.

[0065] Two nucleic acid sequences or polypeptide sequences are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection.

When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, *e.g.*, the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

[0066] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0067] A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

[0068] Algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschul *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (NCBI) web site. The algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits acts as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word size (W) of 28, an expectation (E) of 10, M=1, N=-2, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word size (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

[0069] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.01, more preferably less than about 10^{-5}, and most preferably less than about 10^{-20}.

[0070] Nucleic acid or protein sequences that are substantially identical to a reference sequence include "conservatively modified variants." With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

[0071] As to amino acid sequences, one of skill will recognize that individual substitutions, in a nucleic acid, peptide, polypeptide, or protein sequence which alters a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

[0072] The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(*see, e.g.*, Creighton, *Proteins* (1984)).

[0073] Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other, or a third nucleic acid, under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C. For example, stringent conditions for hybridization, such as RNA-DNA hybridizations in a blotting technique are those which include at least one wash in 0.2X SSC at 55°C for 20 minutes, or equivalent conditions.

[0074] As used herein, the term "promoter" refers to a polynucleotide sequence capable of driving transcription of a DNA sequence in a cell. Thus, promoters used in the polynucleotide constructs of the invention include cis- and trans- acting transcriptional control elements and regulatory sequences that are involved in regulating or modulating the timing and/or rate of transcription of a gene. For example, a promoter can be a cis-acting transcriptional control element, including an enhancer, a promoter, a transcription terminator, an origin of replication, a chromosomal integration sequence, 5' and 3' untranslated regions, or an intronic sequence, which are involved in transcriptional regulation. These cis-acting sequences typically interact with proteins or other biomolecules to carry out (turn on/off, regulate, modulate, *etc.*) gene transcription. Promoters are located 5' to the transcribed gene, and as used herein, include the sequence 5' from the translation start codon (i.e., including the 5' untranslated region of the mRNA, typically comprising 100-200 bp). Most often the core promoter sequences lie within 1-5 kb of the translation start site, more often within 1 kbp and often within 500 bp of the translation start site. By convention, the promoter sequence is usually provided as the sequence on the coding strand of the gene it controls.

[0075] A "constitutive promoter" is one that is capable of initiating transcription in nearly all cell types, whereas a "cell type-specific promoter" initiates transcription only in one or a few particular cell types or groups of cells forming a tissue. In some embodiments, the promoter is secondary cell wall-specific and/or fiber cell-specific. A "fiber cell-specific promoter" refers to a promoter that initiates substantially higher levels of transcription in fiber cells as compared to other non-fiber cells of the plant. A "secondary cell wall-specific promoter" refers to a promoter that initiates substantially higher levels of transcription in cell types that have secondary cell walls, *e.g.*, lignified tissues such as vessels and fibers, which may be found in wood and bark cells of a tree, as well as other parts of plants such as the leaf stalk. In some embodiments, a promoter is fiber cell-specific or secondary cell wall-specific if the transcription levels initiated by the promoter in fiber cells or secondary cell walls, respectively, are at least 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 50-fold, 100-fold, 500-fold, 1000-fold higher or more as compared to the transcription levels initiated by the promoter in other tissues, resulting in the encoded protein substantially localized in plant cells that possess fiber cells or secondary cell wall, *e.g.*, the stem of a plant. Non-limiting examples of fiber cell and/or secondary cell wall specific promoters include the promoters directing expression of the genes IRX1, IRX3, IRX5, IRX7, IRX8, IRX9, IRX10, IRX14, NST1, NST2, NST3, MYB46, MYB58, MYB63, MYB83, MYB85, MYB103, PAL1, PAL2, C3H, CcOAMT, CCR1, F5H, LAC4, LAC17, CADc, and CADd. *See, e.g.*, Turner et al 1997; Meyer et al 1998; Jones et al 2001; Franke et al 2002; Ha et al 2002;Rohde et al 2004; Chen et al 2005; Stobout et al 2005; Brown et al 2005; Mitsuda et al 2005; Zhong et al 2006; Mitsuda et al 2007; Zhong et al 2007a, 2007b; Zhou et al 2009; Brown et al 2009; McCarthy et al 2009; Ko et al 2009; Wu et al 2010; Berthet et al 2011. In some embodiments, a promoter is substantially identical to a promoter from the lignin biosynthesis pathway (*e.g.*, a promoter for a gene encoding a protein shown in Figure 1). Non-limiting examples of promoter sequences are provided herein as SEQ ID NOs:17-28. A promoter originated from one plant species may be used to direct gene expression in another plant species.

[0076] A polynucleotide is "heterologous" to an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, when a polynucleotide encoding a polypeptide sequence is said to be operably linked to a heterologous promoter, it means that the polynucleotide coding sequence encoding the polypeptide is derived from one species whereas the promoter sequence is derived from another, different species; or, if both are derived from the same species, the coding sequence is not naturally associated with the promoter (*e.g.*, is a genetically engineered coding sequence, e.g., from a different gene in the same species, or an allele from a different ecotype or variety, or a gene that is not naturally expressed in the target tissue).

[0077] The term "operably linked" refers to a functional relationship between two or more polynucleotide (*e.g.*, DNA) segments. Typically, it refers to the functional relationship of a transcriptional regulatory sequence to a transcribed sequence. For example, a promoter or enhancer sequence is operably linked to a DNA or RNA sequence if it stimulates or modulates the transcription of the DNA or RNA sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, i.e., they are cis-acting. However, some transcriptional regulatory sequences, such as enhancers, need not be physically contiguous or located in close proximity to the coding sequences whose transcription they enhance.

[0078] The term "expression cassette" refers to a nucleic acid construct that, when introduced into a host cell, results in transcription and/or translation of an RNA or polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition. In the case of both expression of transgenes and suppression of endogenous genes (*e.g.*, by antisense, RNAi, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only substantially identical to a sequence of the gene from which it was derived. As explained herein, these substantially identical variants are specifically covered by reference to a specific nucleic acid sequence.

[0079] The term "plant," as used herein, refers to whole plants and includes plants of a variety of a ploidy levels, including aneuploid, polyploid, diploid, and haploid. The term "plant part," as used herein, refers to shoot vegetative organs and/or structures (*e.g.*, leaves, stems and tubers), branches, roots, flowers and floral organs (*e.g.*, bracts, sepals, petals, stamens, carpels, anthers), ovules (including egg and central cells), seed (including zygote, embryo, endosperm, and seed coat), fruit (*e.g.*, the mature ovary), seedlings, and plant tissue (*e.g.*, vascular tissue, ground tissue, and the like), as well as individual plant cells, groups of plant cells (*e.g.*, cultured plant cells), protoplasts, plant extracts, and seeds. The class of plants that can be used in the methods of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, and multicellular algae.

[0080] The term "biomass," as used herein, refers to plant material that is processed to provide a product, *e.g.*, a biofuel such as ethanol, or livestock feed, or a cellulose for paper and pulp industry products. Such plant material can include whole plants, or parts of plants, *e.g.*, stems, leaves, branches, shoots, roots, tubers, and the like.

[0081] The term "reduced lignin content" encompasses reduced amount of lignin polymer, reduced amount of either or both of the guaiacyl (G) and/or syringyl (S) lignin units, reduced size of a lignin polymer, *e.g.*, a shorter lignin polymer chain due to a smaller number of monolignols being incorporated into the polymer, a reduced degree of branching of the lignin polymer, or a reduced space filling (also called a reduced pervaded volume). In some embodiments, a reduced lignin polymer can be shown by detecting a decrease in the molecular weight of the polymer or a decrease in the number of monolignols by at least 2%, 5%, 10%, 20%, 25%, 30%, 40%, 50%, or more, when compared to the average lignin molecule in a control plant (*e.g.*, a non-transgenic plant). In some embodiments, reduced lignin content can be shown by detecting a decrease in the number or amount of guaiacyl (G) and/or syringyl (S) lignin units in the plant as compared to a control plant (*e.g.*, a non-transgenic plant). In some embodiments, a plant as described herein has reduced lignin content if the amount of guaiacyl (G) and/or syringyl (S) lignin units in the plant is decreased by at least about 2%, 5%, 10%, 20%, 25%, 30%, 40%, 50% or more, as compared to a control plant. Methods for detecting reduced lignin content are described in detail below.

II. Introduction

[0082] Plant cell walls constitute a polysaccharidic network of cellulose microfibrils and hemicellulose embedded in an aromatic polymer known as lignin. This ramified polymer is mainly composed of three phenylpropanoid-derived phenolics (i.e., monolignols) named *p*-coumaryl, coniferyl, and sinapyl alcohols which represent the *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) lignin units (Boerjan *et al.*, 2003). Monolignols have a C_6C_3 carbon skeleton which consists of a phenyl ring (C_6) and a propane (C_3) side chain. Lignin is crucial for the development of terrestrial plants as it confers recalcitrance to plant cell walls. It also provides mechanical strength for upright growth, confers hydrophobicity to vessels that transport water, and acts as a physical barrier against pathogens that degrade cell walls (Boudet, 2007). Notably, lignin content and composition are finely regulated in response to environmental biotic and abiotic stresses (Moura *et al.*, 2010).

[0083] Economically, lignocellulosic biomass from plant cell walls is widely used as raw material for the production of pulp in paper industry and as ruminant livestock feed. Plant feedstocks also represent a source of fermentable sugars for the production of synthetic molecules such as pharmaceuticals and transportation fuels using engineered microorganisms (Keasling, 2010). However, negative correlations exist between lignin content in plant biomass and pulp yield, forage digestibility, or polysaccharides enzymatic hydrolysis (de Vrije *et al.*, 2002; Reddy et al., 2005; Dien *et al.*, 2006; Chen and Dixon, 2007; Dien *et al.*, 2009; Taboada *et al.*, 2010; Elissetche *et al.*, 2011; Studer *et al.*, 2011). Consequently, reducing lignin recalcitrance in plant feedstocks is a major focus of interest, especially in the lignocellulosic biofuels field for which efficient enzymatic conversion of polysaccharides into monosaccharides is crucial to achieve economically and environmentally sustainable production (Carroll and Somerville, 2009).

[0084] Lignin biosynthesis is well characterized and well conserved across land plants (Weng and Chapple 2010). Genetic modifications such as silencing of genes involved in particular steps of this pathway or its regulation have been employed to reduce lignin content (Simmons *et al.*, 2010; Umezawa, 2010) but this approach often results in undesired phenotypes such as dwarfism, sterility, reduction of plant biomass, and increased susceptibly to environmental stress and pathogens (Bonawitz and Chapple, 2010). These pleiotropic effects are generally the consequences of a loss of secondary cell wall integrity, accumulation of toxic intermediates, constitutive activation of defense responses, or depletion of other phenylpropanoid-derived metabolites which are essential for plant development and defense (Li et al., 2008; Naoumkina et al., 2010, Gallego-Giraldo et al., 2011). Alternatively, changing the recalcitrant structure and physico-chemical properties of lignin can be achieved by modifying its monomer composition. For example, incorporation of coniferyl ferulate into lignin improves enzymatic degradation of cell wall polysaccharides (Grabber et al., 2008). Recently, it has been demonstrated that enrichment in 5-hydroxy-G units and reduction in S units in lignin contribute to enhanced saccharification efficiencies without affecting drastically biomass yields and lignin content (Weng et al., 2010; Dien et al., 2011; Fu et al., 2011).

[0085] The present invention provides an alternative strategy to reduce lignin content (*e.g.*, reducing the amount of *p*-hydroxyphenyl (H), guaiacyl (G) and/or syringyl (S) lignin units, or any combination of H-lignin, G-lignin, and S-lignin units). In this strategy, the plant is engineered to express one or more proteins that diverts or shunts a monolignol precursor from a lignin biosynthesis pathway (*e.g.*, a p-coumaryl alcohol, sinapyl alcohol, and/or coniferyl alcohol biosynthesis pathway) into a competitive pathway. By diverting or shunting the production of monolignol precursors from *p*-hydroxyphenyl (H), guaiacyl (G) and/or syringyl (S) lignin unit production to the production of alternative products (*e.g.*, stilbenes, flavonoids, curcuminoids, or bensalacetones, protocatechuates, aromatic amino acids, vitamins, quinones, or volatile compounds) as described herein, the amount of lignin content or its composition, *e.g.*, in specific cell or tissue types such as in secondary cell wall, can be altered in order to enhance saccharification efficiencies without dramatically affecting biomass yield. The present invention also provides plants that are engineered by the method described herein, as well as a plant cell from such a plant, a seed, flower, leaf, or fruit from such a plant, a plant cell that contains an expression cassette described herein for expressing a protein diverts or shunts a monolignol precursor from a lignin biosynthesis pathway into a competitive pathway, and biomass comprising plant tissue from the plant or part of the plant described herein.

III. Plants Having Reduced Lignin Content

A. Expression of a Protein That Diverts a Monolignol Precursor From a Lignin Biosynthesis Pathway

[0086] In one aspect, the present invention provides a method of engineering a plant having reduced lignin content (*e.g.*, reduced amount of lignin polymers, reduced size of lignin polymers, reduced degree of branching of lignin polymers, or reduced space filling). In some embodiments, the plant has reduced lignin content that is substantially localized to specific cell and/or tissue types in the plant. For example, in some embodiments the plant has reduced lignin content that is substantially localized to secondary cell walls and/or fiber cells. In some embodiments, the method comprises:

introducing into the plant an expression cassette comprising a polynucleotide that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway (*e.g.*, a p-coumaryl alcohol, sinapyl alcohol, and/or coniferyl alcohol biosynthesis pathway) in the plant, and wherein the polynucleotide is operably linked to a heterologous tissue-specific promoter; and

culturing the plant under conditions in which the protein that diverts the monolignol precursor from the lignin biosynthesis pathway (*e.g.*, the p-coumaryl alcohol, sinapyl alcohol, or coniferyl alcohol biosynthesis pathway) is expressed.

[0087] In some embodiments, the gene that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway (*e.g.*, a p-coumaryl alcohol, sinapyl alcohol, and/or coniferyl alcohol biosynthesis pathway) reduces the amount of cytosolic and/or plastidial shikimate that is available for the p-coumaryl alcohol, sinapyl alcohol, or coniferyl alcohol biosynthesis pathway; reduces the amount of cytosolic and/or plastidial phenylalanine that is available for the p-coumaryl alcohol, sinapyl alcohol, or coniferyl alcohol biosynthesis pathway; reduces the amount of cinnamate and/or coumarate that is available for the p-coumaryl alcohol, sinapyl alcohol, or coniferyl alcohol biosynthesis pathway; and/or reduces the amount of coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the p-coumaryl alcohol, sinapyl alcohol, or coniferyl alcohol biosynthesis pathway. In some embodiments, the gene that encodes a protein that diverts a monolignol precursor from a lignin biosynthesis pathway (*e.g.*, a p-coumaryl alcohol, sinapyl alcohol, and/or coniferyl alcohol biosynthesis pathway) activates or potentiates a metabolic pathway that competes with the p-coumaryl alcohol, sinapyl alcohol, or coniferyl alcohol biosynthesis pathway biosynthesis pathway for the use of monolignol precursors, including but not limited to a metabolic pathway selected from a stilbene biosynthesis pathway, a flavonoid biosynthesis pathway, and an anthocyanin biosynthesis pathway.

[0088] An expression cassette as described herein, when introduced into a plant, results in the plant having reduced lignin content (*e.g.,* reduced amount of lignin polymers, reduced size of lignin polymers, reduced degree of branching of lignin polymers, or reduced space filling) that is specifically localized to certain cell and/or tissue types (*e.g.,* specifically localized to secondary cell walls and/or fiber cells), thus reducing cell wall recalcitrance to enzymatic hydrolysis while avoiding defects in plant growth or reductions in biomass yield.

[0089] One of skill in the art will understand that the protein that diverts a monolignol precursor from a lignin biosynthesis pathway that is introduced into the plant by an expression cassette described herein does not have to be identical to the protein sequences described herein (*e.g.,* the protein sequences of SEQ ID NOs:2, 4, 6, 8, 10, 12, or 14). In some embodiments, the protein that is introduced into the plant by an expression cassette is substantially identical (*e.g.,* at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to a protein sequence described herein (*e.g.,* a protein sequence of SEQ ID NOs:2, 4, 6, 8, 10, 12, or 14). In some embodiments, the protein that is introduced into the plant by an expression cassette is a homolog, ortholog, or paralog of a protein that diverts a monolignol precursor from a lignin biosynthesis pathway as described herein (*e.g.,* a protein sequence of SEQ ID NOs:2, 4, 6, 8, 10, 12, or 14).

[0090] Gene and protein sequences for enzymes that divert a monolignol precursor from a lignin biosynthesis pathway are described in the Sequence Listing herein. Additionally, gene and protein sequences for these proteins, and methods for obtaining the genes or proteins, are known and described in the art. One of skill in the art will recognize that these gene or protein sequences known in the art and/or as described herein can be modified to make substantially identical enzymes, *e.g.,* by making conservative substitutions at one or more amino acid residues. One of skill will also recognize that the known sequences provide guidance as to what amino acids may be varied to make a substantially identical enzyme. For example, using an amino acid sequence alignment between two or more protein sequences, one of skill will recognize which amino acid residues are not highly conserved and thus can likely be changed without resulting in a significant effect on the function of the enzyme.

Proteins that Reduce the Amount of Shikimate

[0091] In some embodiments, a protein that diverts a monolignol precursor from a lignin biosynthesis pathway reduces the amount of cytosolic and/or plastidial shikimate that is available for the lignin biosynthesis pathway. Examples of such a protein are shown in Figures 2 and 3. In some embodiments, the protein is an enzyme that modifies a shikimate substrate, *e.g.*, a shikimate kinase or a pentafunctional arom protein. In some embodiments, the protein is an enzyme that utilizes shikimate in the synthesis of another compound (*e.g.*, a protocatechuate, an aromatic amino acid, a vitamin, or a quinone), *e.g.*, a dehydryoshikimate dehydratase.

[0092] Non-limiting examples of a shikimate kinase enzyme are described in Gu *et al.*, *J. Mol. Biol.* 319:779-789 (2002). In some embodiments, the protein is a *Mycobacterium tuberculosis* shikimate kinase (AroK) having the amino acid sequence set forth in SEQ ID NO:2. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:2. In some embodiments, the protein is a homolog of a *Mycobacterium tuberculosis* shikimate kinase (AroK) having the amino acid sequence set forth in SEQ ID NO:2. In some embodiments, a polynucleotide encoding the shikimate kinase comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:1.

[0093] Non-limiting examples of a pentafunctional arom protein are described in Duncan *et al.*, *Biochem. J.* 246:375-386 (1987). In some embodiments, the protein is a *Saccharomyces cerevisiae* pentafunctional arom enzyme (Aro1) having the amino acid sequence set forth in SEQ ID NO:4. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:4. In some embodiments, the protein is a homolog of a *Saccharomyces cerevisiae* pentafunctional arom enzyme (Aro1) having the amino acid sequence set forth in SEQ ID NO:4. In some embodiments, a polynucleotide encoding the pentafunctional arom protein comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least

50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:3.

[0094] Non-limiting examples of a dehydryoshikimate dehydratase are described in Teramoto *et al., Appl. Environ. Microbiol.* 75:3461-3468 (2009) and Hansen *et al., Appl. Environ. Microbiol.* 75:2765-2774 (2009). In some embodiments, the protein is a *Corynebacterium glutamicum* dehydryoshikimate dehydratase (QsuB) having the amino acid sequence set forth in SEQ ID NO:6 or a *Podospora anserina* dehydryoshikimate dehydratase (DsDH) having the amino acid sequence set forth in SEQ ID NO:8. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:6 or SEQ ID NO:8. In some embodiments, the protein is a homolog of a *Corynebacterium glutamicum* dehydryoshikimate dehydratase (QsuB) having the amino acid sequence set forth in SEQ ID NO:6 or a homolog of the *Podospora anserina* dehydryoshikimate dehydratase (DsDH) having the amino acid sequence set forth in SEQ ID NO:8. In some embodiments, a polynucleotide encoding the dehydryoshikimate dehydratase comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:5 or SEQ ID NO:7.

Proteins that Reduce the Amount of Phenylalanine

[0095] In some embodiments, a protein that diverts a monolignol precursor from a lignin biosynthesis pathway reduces the amount of cytosolic and/or plastidial phenylalanine that is available for the lignin biosynthesis pathway. Examples of such a protein are shown in Figures 4 and 5. In some embodiments, the protein is an enzyme that modifies a phenylalanine substrate. In some embodiments, the protein is an enzyme that utilizes phenylalanine in the synthesis of another compound (*e.g.*, a volatile compound), *e.g.*, a phenylacetaldehyde synthase or a phenylalanine aminomutase.

[0096] Non-limiting examples of a phenylacetaldehyde synthase are described in Kaminaga *et al., J. Biol. Chem.* 281:23357-23366 (2006) and in Farhi *et al., Plant Mol. Biol.* 72:235-245 (2010). In some embodiments, the protein is a *Petunia hybrida* phenylacetaldehyde synthase

(PAAS) having the amino acid sequence set forth in SEQ ID NO:10. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:10. In some embodiments, the protein is a homolog of a *Petunia hybrida* phenylacetaldehyde synthase (PAAS) having the amino acid sequence set forth in SEQ ID NO:10. In some embodiments, a polynucleotide encoding the phenylacetaldehyde synthase comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:9.

[0097] Non-limiting examples of a phenylalanine aminomutase are described in Feng *et al.,Biochemistry* 50:2919-2930 (2011). In some embodiments, the protein is a *T. canadensis* phenylalanine aminomutase (PAM) having the amino acid sequence set forth in SEQ ID NO:29. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:29. In some embodiments, the protein is a homolog of a *T. canadensis* phenylalanine aminomutase (PAM) having the amino acid sequence set forth in SEQ ID NO:29.

Proteins that Reduce the Amount of Cinnamate and/or Coumarate

[0098] In some embodiments, a protein that diverts a monolignol precursor from a lignin biosynthesis pathway reduces the amount of cinnamate and/or coumarate that is available for the lignin biosynthesis pathway. Examples of such a protein are shown in Figures 6 and 7. In some embodiments, the protein is an enzyme that modifies a cinnamate and/or coumarate substrate, *e.g.*, a cinnamate/p-coumarate carboxyl methyltransferase. In some embodiments, the protein is an enzyme that utilizes cinnamate and/or coumarate in the synthesis of another compound (*e.g.*, a volatile compound, *e.g.*, styrene or p-hydroxystyrene), e.g., phenylacrylic acid decarboxylase or ferulic acid decarboxylase.

[0099] Non-limiting examples of a cinnamate/p-coumarate carboxyl methyltransferase enzyme are described in Kapteyn *et al., Plant Cell* 19:3212-3229 (2007). In some embodiments, the protein is a *Ocimum basilicum* cinnamate/p-coumarate carboxyl methyltransferase (CCMT) having the amino acid sequence set forth in SEQ ID NO:12. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:12. In some embodiments, the protein is a homolog of a *Ocimum basilicum* cinnamate/p-coumarate carboxyl methyltransferase (CCMT) having the amino acid sequence set forth in SEQ ID NO:12. In some embodiments, a polynucleotide encoding the cinnamate/p-coumarate carboxyl methyltransferase comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:11.

[0100] Non-limiting examples of a phenylacrylic acid decarboxylase are described in McKenna *et al., Metab Eng* 13:544-554 (2011). In some embodiments, the protein is a *P. penosaceus* phenylacrylic aicd decarboxylase (PDC) having the amino acid sequence set forth in SEQ ID NO:30. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:30. In some embodiments, the protein is a homolog of a *P. penosaceus* phenylacrylic acid decarboxylase (PDC) having the amino acid sequence set forth in SEQ ID NO:30.

<u>Proteins that Reduce the Amount of Coumaroyl-CoA, Caffeoyl-CoA, and/or Feruloyl-CoA</u>

[0101] In some embodiments, a protein that diverts a monolignol precursor from a lignin biosynthesis pathway reduces the amount of coumaroyl-CoA and/or feruloyl-CoA that is available for the lignin biosynthesis pathway. Examples of such a protein are shown in Figures 8-11. In some embodiments, the protein is an enzyme that modifies a coumaroyl-CoA and/or feruloyl-CoA substrate. In some embodiments, the protein is an enzyme that utilizes coumaroyl-CoA and/or feruloyl-CoA in the synthesis of another compound (*e.g.*, umbelliferone, a volatile compound, scopoletin, chalcone, trihydroxychalcone, stilbene, curuminoid, or benzylacetone), *e.g.*, 2-oxoglutarase-dependent dioxygenase, chalcone synthase, stilbene synthase, cucuminoid synthase, or benzalacetone synthase.

[0102] A non-limiting example of a 2-oxoglutarase-dependent dioxygenase enzyme is described in Vialart *et al., Plant J.* 70:460-470 (2012). In some embodiments, the protein is a *Ruta graveolens* 2-oxoglutarase-dependent dioxygenase (C2'H) having the amino acid sequence set forth in SEQ ID NO:14. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:14. In some embodiments, the protein is a homolog of a *Ruta graveolens* 2-oxoglutarase-dependent dioxygenase (C2'H) having the amino acid sequence set forth in SEQ ID NO:14. In some embodiments, a polynucleotide encoding the oxoglutarase-dependent dioxygenase comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:13.

[0103] Other non-limiting examples of proteins that reduce the amount of coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the lignin biosynthesis pathway chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone synthase (BAS), described in Katsuyama *et al., J. Biol. Chem.* 282:37702-37709 (2007); Sydor *et al., Appl. Environ. Microbiol.* 76:3361-3363 (2010); Jiang *et al., Phytochemistry* 67:2531-2540 (2006); Abe and Morita, *Nat. Prod. Rep.* 27:809 (2010); Dao *et al., Phytochem Rev.* 10:397-412 (2011); Suh *et al., Biochem J.* 350:229-235 (2000); Tropf *et al., J. Biol. Chem.* 270:7922-7928 (1995); Knogge *et al., Arch. Biochem. Biophys.* 250:364-372 (1986); Ferrer *et al., Nat. Struct. Biol.* 6:775-784 (1999); Miyazono *et al., Proteins* 79:669-673 (2010); and Abe *et al., Eur. J. Biochem.* 268:3354-3359 (2001). In some embodiments, the protein is a *Physcomitrella patens* CHS having the amino acid sequence set forth in SEQ ID NO:31; an *Arabidopsis thaliana* CHS having the amino acid sequence set forth in SEQ ID NO:32; a *Vitis vinifera* SPS having the amino acid sequence set forth in SEQ ID NO:33; an *Oryza sativa* CUS having the amino acid sequence set forth in SEQ ID NO:34 or SEQ ID NO:35; or a *Rheum palmatum* BAS having the amino acid sequence set forth in SEQ ID NO:36; or a homolog thereof. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of any of SEQ ID NOs:31, 32, 33, 34, 35, or 36.

Proteins that Activate a Competitive Metabolic Pathway

[0104] In some embodiments, a protein that diverts a monolignol precursor from a lignin biosynthesis pathway activates, upregulates, or potentiates a metabolic pathway that competes with the lignin biosynthesis pathway biosynthesis pathway for the use of monolignol precursors. Non-limiting examples of metabolic pathways that are competitive with the lignin biosynthesis pathway include the stilbene biosynthesis pathway, the flavonoid biosynthesis pathway, the curcuminoid biosynthesis pathway, and the bensalacetone biosynthesis pathway. Thus, in some embodiments, the protein that diverts a monolignol precursor from a lignin biosynthesis pathway is a protein (*e.g.,* a transcription factor, a TALE-based artificial transcription factor (*see* Zhang *et al., Nat. Biotechnol.* 29:149-153 (2011)), or an enzyme) that activates, upregulates, induces, or potentiates a stilbene biosynthesis pathway, a flavonoid biosynthesis pathway, a curcuminoid biosynthesis pathway, or a bensalacetone biosynthesis pathway

[0105] As one non-limiting example, a protein can be expressed that activates, upregulates, induces, or potentiates a flavonoid biosynthesis pathway. The flavonoid biosynthesis pathway utilizes monolignol precursors such as coumaroyl-CoA, caffeoyl-CoA, and feruloyl-CoA from the lignin biosynthesis pathway for the synthesis of flavonoids such as chalcones, flavonones, dihydroflavonols, flavonols, and anthocyanins. *See* Figures 9 and 11. In some embodiments, the protein that diverts a monolignol precursor from a lignin biosynthesis pathway is a protein that activates, upregulates, induces, or potentiates the expression and/or activity of an enzyme in the flavonoid biosynthesis pathway (*e.g.,* an enzyme such as chalcone synthase or flavonol synthase). In some embodiments, the protein that diverts a monolignol precursor from a lignin biosynthesis pathway is a transcription factor. Transcription factors in the flavonoid biosynthesis pathway are known in the art. *See, e.g.,* Bovy *et al., Plant Cell* 14:2509-2526 (2002); Tohge *et al., Plant J.* 42:218-235 (2005); Peel *et al., Plant J.* 59:136-149 (2009); *Pattanaik et al., Planta* 231:1061-1076 (2010); and Hichri *et al., J Exp Botany* 62:2465-2483 (2011). Non-limiting examples of transcription factors in the flavonoid biosynthesis pathway include MYB transcription factors, basic helix-loop-helix (bHLH) transcription factors, and WD40 transcription factors. In some embodiments, the protein is an *Arabidopsis thaliana* PAP1 R2R3 MYB transcription factor having the amino acid sequence set forth in SEQ ID NO:37; an *Arabidopsis thaliana* PAP2 R2R3 MYB transcription factor having the amino acid

Date Reçue/Date Received 2020-05-29

sequence set forth in SEQ ID NO:38; an *Arabidopsis thaliana* TT2 R2R3 MYB transcription factor having the amino acid sequence set forth in SEQ ID NO:39; a *Nicotiana tabacum* NtAn2 R2R3 MYB transcription factor having the amino acid sequence set forth in SEQ ID NO:40; a *Medicago truncatula* LAP1 R2R3 MYB transcription factor having the amino acid sequence set forth in SEQ ID NO:41; a *Zea mays* MYB-C R2R3 transcription factor having the amino acid sequence set forth in SEQ ID NO:42; a *Zea mays* MYC-Lc BHLH transcription factor having the amino acid sequence set forth in SEQ ID NO:43; an *Arabidopsis thaliana* TT8 BHLH transcription factor having the amino acid sequence set forth in SEQ ID NO:44; or a *Vitis vinifera* Myc1 BHLH transcription factor having the amino acid sequence set forth in SEQ ID NO:45; or a homolog thereof. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of any of SEQ ID NOs:37, 38, 39, 40, 41, 42, 43, 44, or 45.

[0106] In some embodiments, a plant is engineered to express two, three, four or more proteins as described herein. In some embodiments, the plant expresses two or more proteins, each of which is identical or substantially identical to SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 42, 43, 44, or 45. In some embodiments, the two or more proteins utilize different substrates or activate different pathways; for example, in some embodiments the plant expresses a first protein that reduces the amount of shikimate that is available for the lignin biosynthesis pathway and a second protein that reduces the amount of phenylalanine that is available for the lignin biosynthesis pathway. In some embodiments, the two or more proteins potentiate or activate the same pathway; for example, in some embodiments the plant expresses a first transcription factor and a second transcription factor that function cooperatively to induce the flavonoid biosynthesis pathway.

Proteins that Produce a Competitive Inhibitor of HCT

[0107] In some embodiments, a plant having reduced lignin content is engineered by expressing or overexpressing a competitive inhibitor of a lignin biosynthesis pathway enzyme (e.g., a molecule that competes with *p*-coumaroyl-CoA and/or shikimate as a substrate for hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT)). In some embodiments, the method comprises:

introducing into the plant an expression cassette comprising a polynucleotide that encodes a protein that produces a competitive inhibitor of hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) in the plant, wherein the polynucleotide is operably linked to a heterologous promoter; and

culturing the plant under conditions in which the protein that produces a competitive inhibitor of HCT is expressed.

[0108] In some embodiments, the protein directly or indirectly produces one or more of the competitive inhibitors protocatechuate, gentisate, catechol, 2,3-dihydroxybenzoate, 3,6-dihydroxybenzoate, or 3-hydroxy-2-aminobenzoate (*e.g.*, by catalyzing the formation of the competitive inhibitor or by catalyzing the formation of a precursor to the competitive inhibitor). Examples of pathways to produce competitive inhibitors of HCT are shown in Figure 27.

[0109] As a non-limiting example, in some embodiments, the competitive inhibitor of HCT is protocatechuate. As shown in Figure 27, protocatechuate can be produced by the enzyme dehydroshikimate dehydratase (QsuB) or by the enzyme dehydroshikimate dehydratase (DsDH). In some embodiments, the protein that produces a competitive inhibitor of HCT is a *Corynebacterium glutamicum* dehydryoshikimate dehydratase (QsuB) having the amino acid sequence set forth in SEQ ID NO:6 or a *Podospora anserina* dehydryoshikimate dehydratase (DsDH) having the amino acid sequence set forth in SEQ ID NO:8. In some embodiments, the protein is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to the amino acid sequence of SEQ ID NO:6 or SEQ ID NO:8. In some embodiments, the protein is a homolog of a *Corynebacterium glutamicum* dehydryoshikimate dehydratase (QsuB) having the amino acid sequence set forth in SEQ ID NO:6 or a homolog of the *Podospora anserina* dehydryoshikimate dehydratase (DsDH) having the amino acid sequence set forth in SEQ ID NO:8. In some embodiments, a polynucleotide encoding the dehydryoshikimate dehydratase comprises a polynucleotide sequence that is identical or substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to SEQ ID NO:5 or SEQ ID NO:7.

B. Plastidial Expression of Proteins

[0110] In some embodiments, the protein that diverts a monolignol precursor from a lignin biosynthesis pathway as described herein is expressed in one or more specific organelles of the plant, *e.g.,* in the plastid of the plant. The polynucleotide sequence encoding the protein that diverts a monolignol precursor from a lignin biosynthesis pathway (*e.g.,* a polynucleotide encoding shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), dehydroshikimate dehydratase (QsuB), phenylacetaldehyde synthase (PAAS), or phenylalanine aminomutase (PAM), *e.g.,* a polynucleotide comprising a sequence that is identical or substantially identical to a polynucleotide sequence of SEQ ID NO:1, 3, 5, 7, or 9, or a polynucleotide comprising a sequence that encodes a polypeptide is identical or substantially identical to an amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, or 29) can be engineered to include a sequence that encodes a targeting or transit signal for the organelle, *e.g.,* a targeting or transit signal for the plastid. Targeting or transit signals act by facilitating transport of proteins through intracellular membranes, *e.g.,* vacuole, vesicle, plastid, and mitochondrial membranes.

[0111] In some embodiments, the plastid targeting signal is a targeting signal described in US Patent No. 5, 510,471. In some embodiments, the plastid targeting signal is identical or substantially identical (*e.g.,* at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to an amino acid sequence of SEQ ID NO:16. In some embodiments, the plastid targeting signal is identical or substantially identical (*e.g.,* at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to a polynucleotide sequence of SEQ ID NO:15. In some embodiments, the organelle targeting signal (*e.g.,* the plastid targeting signal) is linked in-frame with the coding sequence for the protein that diverts a monolignol precursor from a lignin biosynthesis pathway.

C. Promoters

[0112] In some embodiments, the polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway, or the protein that produces a competitive inhibitor of HCT, is operably linked to a heterologous promoter. In some embodiments, the promoter is a cell- or tissue-specific promoter as described below. In some

Date Reçue/Date Received 2020-05-29

embodiments, the promoter is from a gene in the lignin biosynthesis pathway (*e.g.*, a promoter from a gene expressed in the pathway shown in Figure 1). In some embodiments, the promoter is from a gene in the lignin biosynthesis pathway, with the proviso that the promoter is not the native promoter of the polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway or the native promoter of the polynucleotide encoding the protein that produces a competitive inhibitor of HCT to be expressed in the plant. In some embodiments, the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter. In some embodiments, the promoter is identical or substantially identical to a polynucleotide sequence of any of SEQ ID NOs:18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or 28.

Cell- or Tissue-Specific Promoters

[0113] In some embodiments, the polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway, or the protein that produces a competitive inhibitor of HCT, is operably linked to a tissue-specific or cell-specific promoter. In some embodiments, the promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter. The secondary cell wall-specific promoter is heterologous to the polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway, *e.g.*, the promoter and the promoter coding sequence are derived from two different species. A promoter is suitable for use as a secondary cell wall-specific promoter if the promoter is expressed strongly in the secondary cell wall, *e.g.*, in vessel and fiber cells of the plant, but is expressed at a much lower level or not expressed in cells without the secondary cell wall. A promoter is suitable for use as a fiber cell-specific promoter if the promoter is expressed strongly in fiber cells as compared to other non-fiber cells of the plant.

[0114] In some embodiments, the promoter is an IRX5 promoter. IRX5 is a gene encoding a secondary cell wall cellulose synthase *Cesa4 / IRX5*, (Genbank Accession No. AF458083_1). In some embodiments, the promoter is identical or substantially identical to the pIRX5 polynucleotide sequence of SEQ ID NO:17.

[0115] Secondary cell wall-specific promoters are also described in the art. See, for example, Mitsuda *et al., Plant Cell* 17:2993-3006 (2005); Mitsuda *et al., Plant Cell* 19:270-280 (2007); and Ohtani *et al., Plant Journal* 67:499-512 (2011).

[0116] It will be appreciated by one of skill in the art that a promoter region can tolerate considerable variation without diminution of activity. Thus, in some embodiments, a promoter (*e.g.*, a promoter from the lignin biosynthesis pathway, a secondary cell wall-specific promoter, or a fiber cell-specific promoter) is substantially identical (*e.g.*, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical) to a polynucleotide sequence of any of SEQ ID NOs:17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or 28. The effectiveness of a promoter may be confirmed using a reporter gene (*e.g.*, β-glucuronidase or GUS) assay known in the art.

D. Preparation of Recombinant Expression Vectors

[0117] Once the promoter sequence and the coding sequence for the gene of interest (*e.g.*, coding for a protein that diverts a monolignol precursor from the lignin biosynthesis pathway) are obtained, the sequences can be used to prepare an expression cassette for expressing the gene of interest in a transgenic plant. Typically, plant transformation vectors include one or more cloned plant coding sequences (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors may also contain a promoter (*e.g.*, a secondary cell wall-specific promoter or fiber cell-specific promoter as described herein), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

[0118] The plant expression vectors may include RNA processing signals that may be positioned within, upstream, or downstream of the coding sequence. In addition, the expression vectors may include regulatory sequences from the 3'-untranslated region of plant genes, *e.g.*, a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

[0119] Plant expression vectors routinely also include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic resistance genes (*e.g.*, resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin), herbicide resistance genes (*e.g.*, phosphinothricin acetyltransferase), and genes encoding positive selection enzymes (*e.g.* mannose isomerase).

[0120] Once an expression cassette comprising a polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway and operably linked to a promoter as described herein has been constructed, standard techniques may be used to introduce the polynucleotide into a plant in order to modify gene expression. *See, e.g.*, protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture--Crop Species. Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

[0121] Transformation and regeneration of plants are known in the art, and the selection of the most appropriate transformation technique will be determined by the practitioner. Suitable methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumeficiens* mediated transformation. Transformation means introducing a nucleotide sequence in a plant in a manner to cause stable or transient expression of the sequence. Examples of these methods in various plants include: U.S. Pat. Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

[0122] Following transformation, plants can be selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants or the ability to grow on a specific substrate, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic, herbicide, or substrate.

[0123] The polynucleotides coding for a protein that diverts a monolignol precursor from the lignin biosynthesis pathway, as well as the polynucleotides comprising promoter sequences for secondary cell wall-specific promoters or fiber cell-specific promoters, can be obtained according to any method known in the art. Such methods can involve amplification reactions such as PCR and other hybridization-based reactions or can be directly synthesized.

E. Plants in Which Lignin Content Can Be Reduced

[0124] An expression cassette comprising a polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway and operably linked to a promoter, or comprising a polynucleotide encoding the protein that produces a competitive inhibitor of HCT and operably linked to a promoter, as described herein, can be expressed in various kinds of plants. The plant may be a monocotyledonous plant or a dicotyledonous plant. In some embodiments of the invention, the plant is a green field plant. In some embodiments, the plant is a gymnosperm or conifer.

[0125] In some embodiments, the plant is a plant that is suitable for generating biomass. Examples of suitable plants include, but are not limited to, Arabidopsis, poplar, eucalyptus, rice, corn, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, Jatropha, and Brachypodium.

[0126] In some embodiments, the plant into which the expression cassette is introduced is the same species of plant as the promoter and/or as the polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway or encoding the protein that produces a competitive inhibitor of HCT (*e.g.*, a polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway and a secondary cell wall-specific or fiber cell-specific promoter from Arabidopsis is expressed in an Arabidopsis plant). In some embodiments, the plant into which the expression cassette is introduced is a different species of plant than the promoter and/or than the polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway (*e.g.*, a polynucleotide encoding the protein that diverts a monolignol precursor from the lignin biosynthesis pathway and/or a secondary cell wall-specific or fiber cell-specific promoter from Arabidopsis is expressed in a poplar plant). *See, e.g.*, McCarthy *et al.*, *Plant Cell Physiol.* 51:1084-90 (2010); and Zhong *et al.*, *Plant Physiol.* 152:1044-55 (2010).

F. Screening for Plants Having Reduced Lignin Content

[0127] After transformed plants are selected, the plants or parts of the plants can be evaluated to determine whether expression of the protein that diverts a monolignol precursor from the lignin biosynthesis pathway, or expression of the protein that produces a competitive inhibitor of HCT, *e.g.*, under the control of a secondary cell wall-specific promoter or a fiber cell-specific promoter, can be detected, e.g., by evaluating the level of RNA or protein, by measuring enzymatic activity of the protein, and/or by evaluating the size, molecular weight, content, or degree of branching in the lignin molecules found in the plants. These analyses can be performed using any number of methods known in the art.

[0128] In some embodiments, plants are screened by evaluating the level of RNA or protein. Methods of measuring RNA expression are known in the art and include, for example, PCR, northern analysis, reverse-transcriptase polymerase chain reaction (RT-PCR), and microarrays. Methods of measuring protein levels are also known in the art and include, for example, mass spectroscopy or antibody-based techniques such as ELISA, Western blotting, flow cytometry, immunofluorescence, and immunohistochemistry.

[0129] In some embodiments, plants are screened by assessing for activity of the protein being expressed, and also by evaluating lignin size and composition. Enzymatic assays for the proteins described herein (*e.g.,* shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), dehydroshikimate dehydratase (QsuB), phenylacetaldehyde synthase (PAAS), phenylalanine aminomutase (PAM), p-coumarate/cinnamate carboxylmethltransferase (CCMT1), ferulic acid decarboxylase (FDC1), phenylacrylic acid decarboxylase (PDC1), 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone (BAS)) are well known in the art. Lignin molecules can be assessed, for example, by nuclear magnetic resonance (NMR), spectrophotometry, microscopy, klason lignin assays, thioacidolysis, acetyl-bromide reagent or by histochemical staining (*e.g.,* with phloroglucinol).

[0130] As a non-limiting example, any of several methods known in the art can be used for quantification and/or composition analysis of lignin in a plant or plant part as described herein. Lignin content can be determined from extract free cell wall residues using acetyl bromide or Klason methods. *See, e.g.,* Eudes *et al., Plant Biotech. J.* 10:609-620 (2012); Yang *et al., Plant Biotech. J.* (2013) (in press); and Dence *et al.* (eds) *Lignin determination.* Berlin: SpringerVerlag (1992). Extract free cell wall residues correspond to raw biomass, which has been extensively washed to remove the ethanol soluble component. Eudes *et al., Plant Biotech. J.* 10:609-620 (2012); Yang *et al., Plant Biotech. J.* (2013) (in press); Sluiter *et al.,* Determination of structural carbohydrates and lignin in biomass. In: *Laboratory Analytical Procedure*. National Renewable Energy Laboratory, Golden, CO, USA; and Kim *et al., Bio. Res.* 1:56-66 (2008). Lignin composition analysis and G/S lignin subunit determination can be performed using any of various techniques known in the art such as 2D 13C-H1 HSQC NMR spectroscopy (Kim and Ralph, *Org. Biomol. Chem.* 8:576-591 (2010); Kim *et al., Bio. Res.* 1:56-66 (2008)); thioacidolysis method (Lapierre *et al., Plant Physiol.* 119:153-164 (1999); Lapierre et al., Res. Chem. Intermed. 21:397-412 (1995); Eudes *et al., Plant Biotech. J.* 10:609-620 (2012)); derivatization followed by reductive cleavage method (DFRC method; Lu and Ralph, *J. Agr. Food Chem* 46:547-552 (1998) and Lu and Ralph, *J. Agr. Food Chem* 45:2590-2592 (1997)) and pyrolysis-gas chromatograph method (Py-GC method; Sonoda *et al., Anal. Chem.* 73:5429-5435 (2001)) directly from extract free cell wall residues or from cellulolytic enzyme lignin (CEL lignin). CEL lignin derives from cell wall residues, which were hydrolyzed with crude cellulases to deplete the polysaccharide fraction and enrich the lignin one (Eudes *et al., Plant Biotech. J.* 10:609-620 (2012)).

Date Reçue/Date Received 2020-05-29

IV. Methods of Using Plants Having Reduced Lignin Content

[0131] Plants, parts of plants, or plant biomass material from plants having reduced lignification due to the expression of a protein that diverts a monolignol precursor from the lignin biosynthesis pathway or due to the expression of a protein that produces a competitive inhibitor of HCT, *e.g.*, under the control of a secondary cell wall-specific promoter or a fiber cell-specific promoter, can be used for a variety of methods. In some embodiments, the plants, parts of plants, or plant biomass material generate less recalcitrant biomass for use in a conversion reaction as compared to wild-type plants. In some embodiments, the plants, parts of plants, or plant biomass material are used in a saccharification reaction, *e.g.*, enzymatic saccharification, to generate soluble sugars at an increased level of efficiency as compared to wild-type plants. In some embodiments, the plants, parts of plants, or plant biomass material are used to increase biomass yield or simplify downstream processing for wood industries (such as paper, pulping, and construction) as compared to wild-type plants. In some embodiments, the plants, parts of plants, or plant biomass material are used to increase the quality of wood for construction purposes. In some embodiments the plants, parts of plants, or plant biomass material can be used in a combustion reaction, gasification, pyrolysis, or polysaccharide hydrolysis (enzymatic or chemical). In some embodiments, the plants, parts of plants, or plant biomass material are used as feed for animals (*e.g.*, ruminants).

[0132] Methods of conversion, for example biomass gasification, are known in the art. Briefly, in gasification plants or plant biomass material (*e.g.*, leaves and stems) are ground into small particles and enter the gasifier along with a controlled amount of air or oxygen and steam. The heat and pressure of the reaction break apart the chemical bonds of the biomass, forming syngas, which is subsequently cleaned to remove impurities such as sulfur, mercury, particulates, and trace materials. Syngas can then be converted to products such as ethanol or other biofuels.

[0133] Methods of enzymatic saccharification are also known in the art. Briefly, plants or plant biomass material (*e.g.*, leaves and stems) are optionally pre-treated with hot water, dilute alkaline, AFEX (Ammonia Fiber Explosion), ionic liquid or dilute acid, followed by enzymatic saccharification using a mixture of cell wall hydrolytic enzymes (such as hemicellulases, cellulases and beta-glucosidases) in buffer and incubation of the plants or plant biomass material with the enzymatic mixture. Following incubation, the yield of the saccharification reaction can be readily determined by measuring the amount of reducing sugar released, using a standard method for sugar detection, *e.g.* the dinitrosalicylic acid method well known to those skilled in the art. Plants engineered in accordance with the invention provide a higher saccharificaton efficiency as compared to wild-type plants, while the plants' growth, development, or disease resistance is not negatively impacted.

EXAMPLES

[0134] The following examples are provided to illustrate, but not limited the claimed invention.

Example 1: Strategies for Diverting a Monolignol Precursor from the Lignin Biosynthesis Pathway

[0135] The engineered plants of the present invention express one or more genes encoding a protein that diverts a precursor component from the lignin biosynthesis pathway (Figure 1) to a competitive pathway. This diversion reduces the amount of lignin that is produced and increases the amount of product produced by the competitive pathway.

[0136] Figures 2-11 provide exemplary strategies for diverting a precursor component from the lignin biosynthesis pathway. In one strategy (Figures 2 and 3), the monolignol precursor shikimate can be reduced or depleted. For example, the amount of cytosolic and/or plastidial shikimate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a shikimate kinase such as *M. tuberculosis* shikimate kinase ("MtAroK"), a pentafunctional arom protein such as *S. cerevisiae* pentafunctional arom protein ("ScAro1"), a dehydroshikimate dehydratase such as *C. glutamicum* dehydroshikimate dehydratase ("CgQsuB"), or a *P. anserina* dehydroshikimate dehydratase ("PaDsDH").

[0137] In another strategy (Figures 4 and 5), the monolignol precursor phenylalanine can be reduced or depleted. For example, the amount of cytosolic and/or plastidial phenylalanine that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a phenylacetaldehyde such as *P.hybrida* phenylacetaldehyde synthase ("PhPAAS") or a phenylalanine aminomutase such as *T. canadensis* phenylalanine aminomutase ("TcPAM").

[0138] In another strategy (Figures 6 and 7), the monolignol precursors cinnamate and/or p-coumarate are reduced or depleted. For example, the amount of cytosolic cinnamate and/or p-coumarate that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a cinnamate/p-coumarate carboxyl methyltransferase such as *O. basilicum* cinnamate/p-coumarate carboxyl methyltransferase ("ObCCMT1") or a phenylacrylic acid decarboxylase such as *P. pentosaceus* phenylacrylic decarboxylase ("PDC").

[0139] In another strategy (Figures 8-11), the monolignol precursors coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA are reduced or depleted. For example, the amount of cytosolic coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the lignin biosynthesis pathway can be reduced or depleted by expressing a 2-oxoglutarate-dependent dioxygenase such as *R. graveolens* C2'H (2-oxoglutarate-dependent dioxygenase) ("RbC2'H"), a chalcone synthase (CHS), a stilbene synthase (SPS), a cucuminoid synthase (CUS), or a benzalacetone (BAS).

Example 2: Generation of Transgenic Lines Expressing QsuB or DsDH in plastids

[0140] The promoter (pC4H) of the lignin C4H gene from *Arabidopsis* was synthesized with flanking SmaI and AvrII restriction sites at the 3' and 5' ends respectively (Genscript). The encoding sequence of the chloroplastic targeting signal peptide sequence (ctss; Patent US 5510471) was codon optimized and synthesized (Genscript), then amplified by PCR and inserted into the AvrII restriction site located at the 5' end of pC4H using In-Fusion cloning (Clontech). The pC4Hctss DNA fusion was then used to replace the IRX5 promoter from pTKan-pIRX5 (Eudes *et al. Plant Biotechnol J* 10, 609–620 (2012)) using Gateway technology (Invitrogen) and to generate a new pTkan-pC4Hctss-GWR3R2 vector. This vector is designed to clone in-frame with the ctss sequence any gene of interest previously cloned into a pDONR221.P3-P2 vector according to the manufacturer instruction (Invitrogen).

[0141] Codon-optimized nucleotide sequences encoding for the dehydroshikimate dehydratases QsuB from *Corynebacterium glutamicum* (accession number A4QB63) and DsDH from *Podospora anserina* (accession number CAD60599) were synthesized for expression in Arabidopsis (Genescript), cloned in *pDONR221.P3-P2* gateway vector according the manufacturer instruction (Invitrogen), and transferred into pTkan-pC4Hctss-GWR3R2 by LR clonase reaction (Invitrogen) to generate the pTKan-pC4Hctss-QsuB and pTKan-pC4Hctss-DsDH binary vectors respectively. The in-frame fusions of ctss with QsuB and DsDH encoding sequences were verified by sequencing.

[0142] Both constructs were introduced independently into WT Arabidopsis plants (ecotype Col0) via *Agrobacterium tumefaciens*-mediated transformation (Bechtold and Pelletier, *Methods Mol Biol* 82:259-266 (1998)) and several independent S-QsuB and S-DsDH lines harboring ctss::QsuB and ctss::DsDH gene fusions respectively were generated.

Results

[0143] Nine independent lines resistant to kanamycin and therefore harboring the pTKan-pC4Hctss-QsuB construct (S-QsuB lines) were selected and analyzed at the T2 generation. These lines express the dehydroshikimate dehydratase QsuB protein from *Corynebacterium glutamicum* fused to a plastid targeting signal peptide to address the QsuB protein in their plastids. At the rosette stage (3-week-old), S-QsuB lines were phenotypically indistinguishable from wild-type (WT) plants (Figure 11). The biomass from dried senesced stems collected from S-QsuB lines and WT plants was used to perform saccharification analysis. As shown on Figure 12, the amount of reducing sugars released from the biomass of all the S-QsuB lines was higher compared to the amount released from WT plants. In particular, using similar amount of cellulolytic enzyme, the S-QsuB lines #1, 4, and 9 showed improved saccharification efficiencies of up to 3.0 fold compared to WT plants (Figure 12). Moreover, the amount of reducing sugars released from the biomass of S-QsuB lines (#1, #4, #9) and WT plants using different loadings of cellulolytic enzyme cocktail was investigated. As shown on Figure 13, the saccharification efficiency was on average 75% higher for the three S-QsuB lines although 10 times less enzyme was used compared to WT biomass. This result shows that much less cellulolytic enzyme is required to release similar amount of sugars from the biomass of S-QsuB lines compared to that of WT plants.

[0144] Alternatively, five independent lines resistant to kanamycin and therefore harboring the pTKan-pC4Hctss-DsDH construct (S-DsDH lines) were selected and analyzed at the T2 generation. These lines express the dehydroshikimate dehydratase DsDH protein from *Podospora anserine* fused to a plastid targeting signal peptide to address the QsuB protein in their plastids. The biomass from dried senesced stems collected from S-DsDH lines and WT plants was used to perform saccharification analysis. As shown on Figure 14, using identical amount of cellulolytic enzyme, the amount of reducing sugars released over time from the biomass of all the S-DsDH lines was higher compared to the amount released from WT plants, representing an improvement of up to 1.4 fold after 72 h of hydrolysis. Similarly to the S-QsuB lines, this result indicates that the biomass of S-DsDH lines is less recalcitrant to polysaccharide enzymatic digestion compared to WT plants.

Example 3: Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency

ABSTRACT

[0145] Lignin confers recalcitrance to plant biomass used as feedstocks in agro-processing industries or as a source of renewable sugars for the production of bioproducts. The metabolic steps for the synthesis of lignin building blocks belong to the shikimate and phenylpropanoid pathways. Genetic engineering efforts to reduce lignin content typically employ gene-knockout or gene-silencing techniques to constitutively repress one of these metabolic pathways. In this study, we report that expression of a 3-dehydroshikimate dehydratase (QsuB from *Corynebacterium glutamicum*) reduces lignin deposition in *Arabidopsis* cell walls. QsuB was targeted to the plastids to convert 3-dehydroshikimate — an intermediate of the shikimate pathway — into protocatechuate. Compared to wild-type plants, lines expressing QsuB contain higher amounts of protocatechuate, cinnamate, *p*-coumarate, *p*-coumaraldehyde, and coumaryl alcohol. 2D-NMR spectroscopy, thioacidolysis, and pyrolysis-gas chromatography/mass spectrometry (pyro-*GC/MS)* reveal an increase of *p*-hydroxyphenyl units and a reduction of guaiacyl units in the lignin of QsuB lines, while size-exclusion chromatography indicates a lower degree of lignin polymerization. Our data show that the expression of QsuB primarily affects one of the key enzymatic steps within the lignin biosynthetic pathway. Finally, biomass from these lines exhibits more than a twofold improvement in saccharification efficiency. We conclude that the expression of QsuB in plants, in combination with specific promoters, is a promising gain-of-function strategy for spatio-temporal reduction of lignin in plant biomass.

SIGNIFICANCE

[0146] Lignin is a complex aromatic polymer found in plant cells walls that is largely responsible for the strength and toughness of wood. These properties also confer "recalcitrance" to biomass, so materials high in lignin content are more difficult to break down in processes such as production of biofuels. Efforts to reduce lignin content through altering plant gene expression often result in reduced biomass yield and compromise plant fitness. In this study, we present an effective alternative strategy: reducing lignin content and biomass recalcitrance through expression of a bacterial 3-dehydroshikimate dehydratase in plants. We demonstrate that this strategy achieved dramatic changes in the lignin composition and structure in transgenic plants, as well as improved conversion of biomass into fermentable sugars.

INTRODUCTION

[0147] Plant cells walls are the primary source of terrestrial biomass and mainly consist of cellulosic and hemicellulosic polysaccharides impregnated with lignins. Lignins are polymers of p-hydroxycinnamyl alcohols (*i.e.,* monolignols), which are synthesized inside the cells, exported to the cell wall, and ultimately undergo oxidative polymerization via laccase and peroxidase activities. The main monolignols — p-coumaryl, coniferyl, and sinapyl alcohols

— give rise to the p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin units, respectively (1). Lignification generally confers mechanical strength and hydrophobicity in tissues that develop secondary cell walls, such as sclerenchyma (*i.e.*, fibers) and xylem vessels. In addition to its essential role for upright growth, lignin also serves as a physical barrier against pathogens that degrade cell walls (2).

[0148] Lignocellulosic biomass is used for pulp and paper manufacture, ruminant livestock feeding, and more recently has been considered an important source of simple sugars for fermentative production of intermediate or specialty chemicals and biofuels (3). It is well-documented that lignin in plant biomass negatively affects pulp yield, forage digestibility, and polysaccharide saccharification (4–6). This has prompted major interest in developing a better understanding of lignin biosynthesis to reduce biomass recalcitrance by modifying lignin content and/or composition.

[0149] The shikimate pathway, which is located in plastids in plants, provides a carbon skeleton for the synthesis of phenylalanine, the precursor of the cytosolic phenylpropanoid pathway responsible for the biosynthesis of monolignols (Fig. 20). All the metabolic steps and corresponding enzymes for both pathways are known and well-conserved across land plants (7–10). Classic approaches to lignin reduction have relied on genetic modifications, such as transcript reduction and allelic variation of specific genes from the phenylpropanoid pathway (11, 12). However, these strategies often result in undesired phenotypes — including dwarfism, sterility, and increased susceptibly to environmental stresses — due to loss of cell-wall integrity, depletion of other phenylpropanoid-related metabolites, accumulation of pathway intermediates, or the constitutive activation of defense responses (13, 14). Such negative effects are unfortunately difficult to avoid because of the non-tissue specificity of the strategies employed: allelic variations are transmitted to every cell of the plant during cell divisions, and small interfering RNAs generated for gene silencing generally move from cell-to-cell and over long distance in vegetative tissues (15).

[0150] Alternatively, there are novel and promising gain-of-function strategies that involve expression of specific proteins to reduce the production of the three main monolignols or change their ratios. Using specific promoters with restricted expression patterns, these strategies would enable the alteration of lignin at later developmental stages or, for example, only in certain tissues such as fibers — without compromising the functionality of conductive vessels for the transport of water (14). Examples of such expressed proteins are transcription factors that act as negative regulators of lignin biosynthesis (16-19); enzymes that use intermediates of the lignin pathway for the synthesis of derived metabolites (20-22);

engineered enzymes that modify monolignols into their non-oxidizable forms (23); or proteins that mediate the post-transcriptional degradation of enzymes from the lignin biosynthetic pathway (24).

[0151] In this study, we report for the first time on the expression of a bacterial 3-dehydroshikimate dehydratase in *Arabidopsis* (25). We selected QsuB from *C. glutamicum* and targeted it to the plastids to convert the shikimate precursor 3-dehydroshikimate into protocatechuate, with the aim of reducing lignin content and modifying its composition and structure in the biomass of transgenic lines. Metabolomic analysis of plants expressing QsuB revealed higher amounts of cinnamate, *p*-coumarate, and of the two direct precursors of H-lignin units: *p*-coumaraldehyde and *p*-coumaryl alcohol. Conversely, the direct precursors of G and S units — coniferaldehyde, coniferyl alcohol, sinapaldehyde, and sinapyl alcohol — were reduced. Lignin content was severely reduced in these transgenic lines and exhibited an enrichment of H units at the expense of G units and a lower polymerization degree. Compared to those of wild-type plants, cell walls from lines expressing QsuB released significantly higher amounts of simple sugars after cellulase treatment and required less enzyme for saccharification. Collectively, these results support the hypothesis that expression of a plastidic QsuB affects one of the enzymatic steps within the lignin biosynthetic pathway.

RESULTS

Targeted expression of QsuB in Arabidopsis

[0152] A sequence encoding QsuB was cloned downstream of the sequence encoding for a plastid-targeting signal peptide (SCHL) for expression in plastids. Using transient expression in tobacco, we first confirmed that QsuB was correctly targeted to the plastids by analyzing its subcellular localization when fused at the C-terminus to a YFP marker (Fig. 21). The *schl-qsuB* sequence was cloned downstream of the *Arabidopsis C4H* promoter for expression in lignifying tissues of *Arabidopsis*. Western blot analysis confirmed that QsuB was expressed in stems of several T2 plants homozygous for the *pC4H::schl::qsuB* construct (Fig. 16). Based on the migration of molecular weight markers, QsuB was detected at around 70 kDa, which corresponds to the theoretical size of its native sequence after cleavage of the chloroplast transit peptide (Fig. 16). Five lines with different QsuB expression levels (*C4H::qsuB-1*, *-3*, *-6*, *-7*, and *-9*) were selected for biomass measurement. Although a height reduction was observed for these lines, only two of them (*C4H::qsuB-1* and *-9*) showed a slight decrease of biomass yield (stem dry weight) by 18% and 21%, respectively (Table 1).

Table 1. Height and dry weight of the main inflorescence stem of senesced mature wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) plants.

Plant line	Height (cm) Mean ± SE	Dry weight (mg) Mean ± SE	n
WT	47.3 ± 0.8	271.0 ± 11.1	24
C4H::qsuB-1	36.6 ± 1.0***	221.3 ± 11.0**	20
C4H::qsuB-3	38.8 ± 0.7***	244.4 ± 13.4	20
C4H::qsuB-6	35.9 ± 0.9***	254.1 ± 12.7	20
C4H::qsuB-7	41.0 ± 0.9***	251.3 ± 17.4	20
C4H::qsuB-9	31.8 ± 0.7***	214.4 ± 14.2**	20

n = number of plants analyzed. Asterisks indicate significant differences from the wild-type using the unpaired Student's t-test (*$P < 0.05$; ** $P < 0.005$; ***$P < 0.001$).

Metabolite analysis of *C4H::qsuB* lines

[0153] Methanol soluble metabolites from stems of the *C4H::qsuB-1* and *C4H::qsuB-9* lines were extracted for analysis (Table 2, Fig. 22). Compared to wild-type plants, protocatechuate content was increased 53- and 485-fold in those two transgenic lines, respectively. However, except for tyrosine in line *C4H::qsuB-9*, no significant reduction was observed for the content of several metabolites derived from the shikimate pathway in plastids such as salicylate and aromatic amino acids. Instead, salicylate was slightly increased, 1.3–1.4-fold, in both lines and phenylalanine was 1.6-fold higher in line *C4H::qsuB-1*. Interestingly, several metabolites from the phenylpropanoid pathway were increased in the transgenic lines. Cinnamate and *p*-coumaraldehyde were detected only in transgenic lines; while *p*-coumarate and *p*-coumaryl alcohol contents were increased, compared to those of wild type, 14–18-fold and 3.5–30-fold, respectively. Kaempferol and quercetin, two flavonols derived from *p*-coumaroyl-CoA, were also found in higher amounts in both *C4H::qsuB* lines. The direct precursors of G- and S-lignin units were negatively altered; coniferaldehyde was reduced ~40% in both transgenic lines, while coniferyl alcohol, sinapaldehyde, and sinapyl alcohol were decreased twofold in *C4H::qsuB-9* (Table 2).

[0154] Cell wall-bound metabolites released from cell wall residues by mild alkaline hydrolysis were also analyzed (Table 3). Protocatechuate was found in cell walls of the *C4H::qsuB* lines but not in those from wild-type plants. The content of *p*-coumarate was significantly increased in line *C4H::qsuB-1*, whereas ferulate was reduced in both transgenic lines.

Table 2. Quantitative analysis of methanol-soluble metabolites in stems from 6-wk-old wild-type (WT) and pC4H::schl::qsuB (C4H::qsuB) plants.

	Mean ± SE

Metabolites	WT	*C4H::qsuB-1*	*C4H::qsuB-9*
Protocatechuate$^{\alpha}$	2.04 ± 0.4	108.0 ± 24.8****	991.9 ± 60.7****
Tryptophan$^{\alpha}$	3.7 ± 0.5	3.4 ± 0.2	3.4 ± 0.2
Phenylalanine$^{\alpha}$	2.9 ± 0.2	4.7 ± 0.2***	3.3 ± 0.2
Tyrosine$^{\alpha}$	5.0 ± 1.1	4.2 ± 0.6	2.7 ± 0.2*
Sinapyl alcohol$^{\alpha}$	4.1 ± 0.3	5.7 ± 0.4**	1.9 ± 0.4***
Quercetin$^{\alpha}$	16.1 ± 3.6	12.8 ± 0.6	24.6 ± 1.8*
Kaempferol$^{\alpha}$	159.4 ± 31.6	239.8 ± 9.7**	260.2 ± 8.8**
p-Coumarate$^{\beta}$	6.8 ± 1.2	123.1 ± 9.9****	93.7 ± 12.8****
p-Coumaryl alcohol$^{\beta}$	7.6 ± 1.9	26.8 ± 4.8**	229.6 ± 32.8****
Coniferyl aldehyde$^{\beta}$	28.6 ± 1.8	18.1 ± 2.3**	16.6 ± 1.8***
Coniferyl alcohol$^{\beta}$	828.5 ± 99.2	671.0 ± 63.2	457.0 ± 62.2**
Sinapyl aldehyde$^{\beta}$	59.2 ± 3.9	68.1 ± 8.7	36.4 ± 3.1***
Salicylate$^{\beta}$	655.3 ± 30.7	854.4 ± 63.1**	905.7 ± 111.5*
Cinnamate$^{\beta}$	nd$^{\varphi}$	977.2 ± 389.1	144.3 ± 50.5

α (µg/g fresh weight)

β (µg/g fresh weight)

φ Using a detection limit of 34 ng/g fresh weight

Values are means of four biological replicates (n = 4). nd, not detected. Asterisks indicate significant differences from the wild type using the unpaired Student's t-test (*P < 0.1; **P < 0.05; ***P < 0.005; ****P < 0.001).

Table 3. Quantitative analysis of cell wall-bound aromatics in stems from extractive-free senesced mature wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) plants.

	Mean ± SE (µg/g dry weight)		
Metabolite	**WT**	***C4H::qsuB-1***	***C4H::qsuB-9***
Protocatechuate	nd	6.3 ± 0.4	6.7 ± 1.4
p-Coumarate	15.8 ± 3.0	32.4 ± 2.5*	20.4 ± 1.0
Ferulate	18.1 ± 0.7	7.8 ± 0.5**	5.3 ± 0.1**

Values are means of four biological replicates (n = 3). nd, not detected. Asterisks indicate significant differences from the wild type using the unpaired Student's t-test (*P < 0.05; **P < 0.005; ***P < 0.001).

Compositional analysis of cell wall from *C4H::qsuB* lines

[0155] Using the Klason method, the lignin content measured in the stem of lines *C4H::qsuB-1* and *C4H::qsuB-9* was reduced 50% and 64%, respectively, compared to that of wild type (Table 4). Analysis of the cell-wall monosaccharide composition showed higher amounts of glucose (+ 4-10%), xylose (+ 13-19%), and other less abundant sugars in the transgenic lines, resulting in 8% increase in total cell-wall sugars for the *C4H::qsuB-1* line and an 11% increase for *C4H::qsuB-9* line (Table 4).

Table 4. Chemical composition of senesced mature stems from wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) plants.

	Mean ± SE (mg/g cell wall)		
Component	**WT**	***C4H::qsuB-1***	***C4H::qsuB-9***
Glucose	376.7 ± 5.0	391.6 ± 2.9*	416.0 ± 0.9**
Xylose	173.0 ± 2.0	199.5 ± 2.2**	212.9 ± 0.2**
Galacturonic acid	60.8 ± 2.0	70.8 ± 0.5*	63.1 ± 0.8
Galactose	20.5 ± 0.5	23.3 ± 0.1*	20.2 ± 0.3
Arabinose	17.1 ± 0.4	19.4 ± 0.1*	16.8 ± 0.3
Rhamnose	12.1 ± 0.3	14.1 ± 0.2**	13.0 ± 0.2
Fucose	1.8 ± 0.1	2.3 ± 0.1	2.0 ± 0.1
Glucuronic acid	7.1 ± 0.1	7.3 ± 0.1	8.2 ± 0.2*
Total sugars	669.1 ± 6.8	728.4 ± 5.1**	752.3 ± 2.8**
Klason lignin	191.5 ± 9.5	96.2 ± 8.0**	68.4 ± 5.8**
Acid soluble lignin	4.5 ± 0.4	5.0 ± 0.7	4.7 ± 0.9

Values are means ± SE of triplicate analyses ($n = 3$). Asterisks indicate significant differences from the wild type using the unpaired Student's t-test (*$P < 0.05$; **$P < 0.005$).

Lignin monomeric composition and structure in *C4H::qsuB* lines

[0156] Determination of the lignin monomer composition, using thioacidolysis, indicated an increase in the relative amount of H units in transgenic lines. H units represented 12.7% and 27.9% of the total lignin monomers in lines *C4H::qsuB-1* and *C4H::qsuB-9*, which corresponds to 21- and 46-fold increases compared to that of wild type, respectively (Table 5). The relative amount of G units in transgenics (~45%) was also reduced compared to wild type (~64%), whereas S units were higher in *C4H::qsuB-1* and lower in *C4H::qsuB-9* (Table 5).

[0157] NMR (2D ^{13}C–^{1}H-correlated, HSQC) spectra of cell-wall material from *C4H::qsuB-1* and *C4H::qsuB-9* lines were also obtained for determination of lignin composition and structure. Analysis of the aromatic region of the spectra confirmed the higher relative amount of H units in both *C4H::qsuB* lines (29% and 64.4% respectively) compared to that in wild type (3.6%), as well as a reduction of G units (Fig. 17). Moreover, analysis of the aliphatic region of the spectra indicated a strong reduction of phenylcoumaran (β-5) and resinol (β-β) linkages in the lignin of the transgenic lines (Fig. 23).

[0158] Finally, cell-wall material from stems of wild-type and *C4H::qsuB* lines were analyzed by pyro-GC/MS. For each line, identification and relative quantification of the pyrolysis products derived from H, G, or S units allowed determination of H/G/S ratios (Figure 28). Compared to wild type, H units were increased 3.5- and 10-fold, and G units were reduced 1.4- and 2.2-fold, in lines *C4H::qsuB-1* and *C4H::qsuB-9*, respectively.

Table 5. Main H, G, and S lignin-derived monomers obtained by thioacidolysis of extractive-free senesced mature stems from wild-type (WT) and *pC4H::schl::qsuB* (*C4H::qsuB*) plants.

	WT	*C4H::qsuB-1*	*C4H::qsuB-9*
Total yield (μmol/g CWR)	263.5 (22.7)	116.3 (11.8)*	73.5 (2.1)**
Total yield (μmol/g KL)	1372.5 (118.5)	1211.8 (122.6)	1081.2 (30.7)*
%H	0.6 (0.03)	12.7 (0.78)**	27. 9 (0.38)**
%G	63.7 (0.46)	46.5 (1.94)*	44.9 (0.40)*
%S	35.7 (0.43)	40.8 (1.16)*	27.2 (0.02)*

Values in parentheses are the SE from duplicate analyses. Asterisks indicate significant differences from the wild type using the unpaired Student's t-test (*$P < 0.05$; **$P < 0.01$).

Lignins from *C4H::qsuB* lines have a lower polymerization degree

[0159] Lignin fractions were isolated from wild-type and *C4H::qsuB* lines for analysis of their polydispersity using size-exclusion chromatography (SEC). Elution profiles acquired by monitoring UV-F fluorescence of the dissolved lignin revealed differences between wild-type and transgenic lines (Fig. 18). The total area of the three mass peaks, corresponding to the largest lignin fragments detected between 7.8 min and 12.5 min, was significantly reduced in *C4H::qsuB* lines compared to wild type. Similarly, intermediate molecular mass material, which elutes in a fourth peak between 12.5 min and 18 min, was also less abundant in *C4H::qsuB* lines. Conversely, the area corresponding to the smallest lignin fragments, detected between 18 min and 23.5 min, was increased in the transgenic lines. These results demonstrate a reduction in the degree of polymerization of lignins purified from plants expressing QsuB compared to that of wild type.

Biomass from *C4H::qsuB* lines shows improved saccharification

[0160] Saccharification assays on stem material were conducted to evaluate the cell-wall recalcitrance of the *C4H::qsuB* lines. As shown in Fig. 19A, higher amounts of sugars were released after 72 hr enzymatic hydrolysis of biomass from the *C4H::qsuB* lines (-1, -3, -6, -7 and -9) compared to those of wild type in all pretreatments tested. Saccharification improvements ranged between 79–130% after hot water; 63–104% after dilute alkali; and 26–40% after dilute acid pretreatments (Fig. 19A). Moreover, similar saccharification experiments using hot water pretreated biomass, at 5x lower cellulase loadings, revealed that biomass from all *C4H::qsuB* lines releases more sugar than that of wild type hydrolyzed with a typical enzyme loading (Fig. 19B). Taken together, these data demonstrate that cellulose from the *C4H::qsuB* lines is less recalcitrant to cellulase digestion and requires a lower amount of enzyme to be converted into high yields of fermentable sugars.

DISCUSSION

[0161] Gain-of-function strategies have several advantages for the manipulation of metabolic pathways. For example, they can be used to bioengineer lignin deposition in plants via better spatio-temporal control of monolignol production in lignifying cells, and to adjust lignin composition and its biophysical properties (26). Therefore, identification of proteins in which *in planta*-expression results in modifications of lignin content or composition is of particular interest and presents novel opportunities. In this work, we demonstrate that expression of the 3-dehydroshikimate dehydratase QsuB in plastids leads to drastic reduction and compositional changes of lignin in *Arabidopsis* (Table 4). As a result, biomass from these transgenic plants exhibits much higher saccharification efficiency after pretreatment (Fig. 19A), which is a highly desired trait for several agro-industries and the bioenergy sector. Moreover, the efficiency of this approach to decrease lignin content in plant biomass allows a reduction of hydrolytic enzyme loadings by at least five-fold, while retaining greater saccharification potential than control plants hydrolyzed at standard enzyme loading (Fig. 19B). Consequently, the transfer of this technology to energy crops should have a great impact on the cost-effectiveness of cellulosic biofuels production, since enzyme cost is the major barrier in this process (27).

[0162] In this study, as a proof of concept, we used the promoter of the *AtC4H* gene to ensure strong QsuB expression in all lignifying tissues of the plant. This resulted in a slight decrease of plant height for all the lines; but no significant reductions in biomass yield except for that of two transgenic lines, which expressed QsuB very strongly (Table 1; Fig. 16) and exhibited — in some stem transverse sections (Fig. 24) — evidence of vessel collapse that could impair xylem conductivity (14). Nevertheless, our strategy offers the potential to overcome these defects by selecting more stringent promoters (e.g., fiber-specific) that would exclude QsuB expression from xylem-conductive elements (26, 28). Moreover, translation of our technology from model plant to crops is expected to be straightforward: it is based solely on the expression of QsuB, does not require any particular genetic backgrounds, and the lignin and shikimate pathways are well-conserved among vascular plants.

[0163] A direct consequence of QsuB expression is the accumulation of protocatechuate in the biomass of transgenic plants (~1% dry weight in line *C4H::qsuB-9*; Table 2). Considering the beneficial properties of protocatechuate in the bio-based polymer industry and human health sector, such de novo production adds extra commercial value to the biomass of plants expressing QsuB (29, 30). Much higher amounts of protocatechuate were recovered after acid treatment of the methanol-soluble extracts from transgenic plants (data not shown), which suggests its conjugation in the cytosol after export from the plastids. Interestingly, QsuB expression did not affect substantially the level of metabolites derived from the shikimate pathway, such as aromatic amino acids and salicylate, suggesting that plastidic 3-dehydroshikimate is not limiting (Table 2). On the other hand, a buildup of cinnamate and *p*-coumarate was observed in these lines, accompanied by an accumulation of *p*-coumaraldehyde and *p*-coumaryl alcohol pools (Table 2 and Fig. 22).

[0164] Analysis of the lignin monomeric composition — using 2D NMR spectroscopy, thioacidolysis, and pyro-*GC/MS* — unequivocally demonstrated an increase in H units in plants expressing QsuB (Fig. 17 and Fig. 28; Table 5). These data could explain the reduced degree of polymerization of these lignins, which has been previously observed in various lignin mutants that exhibit high content of H units, incorporation of which typically slows or stops lignin-chain elongation (31, 32; Fig. 18). Therefore, reduced lignin-polysaccharide crosslinking within the biomass of the transgenic lines is expected, and this could contribute to its superior enzymatic digestibility.

[0165] A low lignin content rich in H-units corresponds to a phenotype previously characterized in plants down-regulated for hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT), *p*-coumarate 3-hydroxylase (C3H), or caffeoyl shikimate esterase (CSE). This suggests that an alteration of these biosynthetic steps has occurred in the *C4H::qsuB* lines (10, 32, 33). A possible explanation is that QsuB activity in plastids affects the export of shikimate from the plastids to the cytosol. This would indirectly limit the availability of cytosolic shikimate used for the enzymatic step catalyzed by HCT. The distribution of shikimate between plastids and the cytosol is still poorly understood, and shikimate levels were below the detection limit in our stem extracts from wild-type and transgenic plants. Alternatively, because previous studies reported a substrate flexibility of HCTs (34, 35), the large accumulation of protochatechuate could act as inhibitor of AtHCT, which couples *p*-coumaroyl-CoA and shikimate. Using an *in vivo* enzymatic assay to determine the substrate preference of AtHCT, we confirmed its affinity for *p*-coumaroyl-CoA and shikimate, but also demonstrated its capacity to accept protocatechuate and several other substrates such as catechol, 3,6-dihydroxybenzoate, 3-hydroxy-2-aminobenzoate, and 2,3-dihydroxybenzoate (Fig. 25). Therefore, we cannot exclude the possibility that the protocatechuate pool accumulated in *C4H::qsuB* plants exerts a competitive inhibition of HCT and limits the synthesis of coumaroyl shikimate required for the production of G- and S-lignin units.

MATERIALS AND METHODS

Plant material and growth conditions

[0166] *Arabidopsis thaliana* (ecotype Columbia, Col-0) seeds were germinated directly on soil. Growing conditions were 150 μmol/m^2/s, 22 °C, 60% humidity, and 10 h of light per day for the first 4-5 wk, followed by 14 h of light per day until senescence. Selection of T1 and T2 transgenic plants was made on Murashige and Skoog vitamin medium (PhytoTechnology Laboratories, Shawnee Mission, KS), supplemented with 1% sucrose, 1.5% agar, and 50 μg/mL kanamycin.

Generation of binary vectors

[0167] The promoter *p35S*, with a single enhancer, was amplified by PCR from pRT100 with phosphorylated primers F-p35S (5'-GTCAACATGGTGGAGCACGACAC-3') and R-p35S (5'-CGAGAATCTAGATTGTCCTCTCCAAATGAAATGAACTTC-3'), and cloned into a *Sma*I-digested dephosphorylated pTkan vector (36) to generate a pTKan-*p35S* vector. Subsequently, a GW-YFP cassette was extracted from the pX-YFP vector (37) by *Xho*I/*Spe*I digestion, and ligated into a *Xho*I/*Spe*I-digested pTKan-*p35S* vector to generate the pTkan-*p35S*-GWR1R2-YFP vector.

[0168] A chimeric DNA construct was synthesized (GenScript, Piscatway, NJ): it was flanked by the gateway sequences attB4r (5'-end) and attB3r (3'-end), and contained, in the following order, the *tG7* terminator; the restriction sites *Sma*I, *Kpn*I, *Hin*dIII and *Xho*I; a 2.9-Kb sequence corresponding to the *Arabidopsis* C4H promoter (*pC4H*); and a sequence encoding a plastid targeting signal (SCHL; 38). This attB4r-*tG7-pC4H-schl*-attB3r construct was then subcloned into the Gateway pDONR221-P4rP3r entry vector by BP recombination (Life technologies, Foster City, CA, USA) to generate pENTR-L4-tG7-*pC4H-schl*-L3. An LR recombination reaction was performed with pTkan-*pIRX5*-GW (21), pENTR-L1-*pLac*-lacZalpha-L4 (Life technologies, Foster City, CA, USA), pENTR-L3-*pLac*-Tet-L2 (Life technologies, Foster City, CA, USA), and pENTR-L4-*tG7-pC4H*::*schl*-L3. The obtained construct was subsequently digested by *Sma*I to remove the *pLac*-lacZalpha and *tG7* fragments. The *pLac*-Tet fragment was replaced by the gateway cassette using BP recombination to generate the pTKan-*pC4H*::*schl*-GWR3R2 vector.

Generation of a pTkan-*pC4H*::*schl-qsuB* plasmid and plant transformation

[0169] A gene sequence encoding QsuB from *C. glutamicum* (GenBank accession number YP_001137362.1) without stop codon and flanked with the Gateway™ *att*B3 (5'-end) and *att*B2 (3'-end) recombination sites was synthesized for expression in *Arabidopsis* (GenScript,

Date Reçue/Date Received 2020-05-29

Piscatway, NJ) and cloned into the Gateway pDONR221-P3P2 entry vector by BP recombination (Life technologies, Foster City, CA, USA). A sequence-verified entry clone was LR recombined with the pTKan-*pC4H::schl*-GWR3R2 vector to generate the pTKan-*pC4H::schl-qsuB* construct, which was introduced into wild-type *Arabidopsis* plants (ecotype Col-0) via *Agrobacterium*-mediated transformation (39).

Western blot analysis

[0170] Proteins from *Arabidopsis* stems were extracted using a buffer containing 250 mM Tris-HCl pH 8.5, 25 mM EDTA, 2 mM DTT, 5 mM β-mercaptoethanol, and 10% sucrose; and were quantified using the Bradford method (40). Proteins (15 μg) were separated by SDS-PAGE, blotted, and immunodetected using a universal antibody, as previously described (41).

Methanol-soluble metabolites extraction

[0171] *Arabidopsis* stems of 6-wk-old wild-type and transgenic lines were collected in liquid nitrogen and stored at -80 °C until further utilization. Prior the metabolite extraction, collected stems were pulverized in liquid nitrogen. For extraction of methanol-soluble metabolites, 700–1,000 mg of frozen stem powder was mixed with 2 ml of 80% (v/v) methanol-water and mixed (1,400 rpm) for 15 min at 70 °C. This step was repeated four times. Pooled extracts were cleared by centrifugation (5 min, 20,000 × ***g***, at room temperature), mixed with 4 mL of analytical grade water and filtered using Amicon™ Ultra centrifugal filters (10,000 Da MW cutoff regenerated cellulose membrane; EMD Millipore™, Billerica, MA). Filtered extracts were lyophilized and the resulting pellets dissolved in 50% (v/v) methanol-water prior to LC-MS analysis. An acid-hydrolysis of the samples was performed for the quantification of protocatechuate, salicylate, and flavonols; an aliquot of the filtered extracts was dried under vacuum, resuspended with 1 N HCl and incubated at 95 °C for 3 h. The mixture was subjected to three ethyl acetate partitioning steps. Ethyl acetate fractions were pooled, dried in vacuo, and resuspended in 50% (v/v) methanol-water prior to LC-MS analysis.

Cell-wall bound aromatics extraction

[0172] Senesced stems were ball-milled using a Mixer Mill™ MM 400 (Retsch Inc., Newtown, PA) and stainless steel balls for 2 min at 30 s^{-1}. Extractive-free cell-wall residues (CWR) were obtained by sequentially washing 60 mg of ball-milled stems with 1 mL of 96% ethanol at 95 °C twice for 30 min and mixing with 1 mL of 70% ethanol twice for 30 sec. The resulting CWR were dried in vacuo overnight at 30 °C. The CWR (6 mg) were mixed with

Date Reçue/Date Received 2020-05-29

500 μL of 2 M NaOH and shaken at 1,400 rpm for 24 h at 30 °C. The mixture was acidified with 100 μL of concentrated HCl, and subjected to three ethyl acetate partitioning steps. Ethyl acetate fractions were pooled, dried in vacuo, and suspended in 50% (v/v) methanol-water prior to LC-MS analysis.

LC-MS analysis

[0001] As previously described in Bokinsky *et al.* (42) and Eudes *et al.* (43) — aromatic amino acids, and aromatic acids and aldehydes, respectively — were analyzed using high-performance liquid chromatography (HPLC), electrospray ionization (ESI), and time-of-flight (TOF) mass spectrometry (MS). Aromatic alcohols were analyzed by HPLC — atmospheric pressure chemical ionization (APCI) — TOF MS. Their separation was conducted on an Agilent 1200 Series Rapid Resolution HPLC system (Agilent™ Technologies Inc., Santa Clara, CA, USA) using a Phenomenex Kinetex™ XB-C18 (100 mm length, 2.1 mm internal diameter, and 2.6 μm particle size; Phenomenex, Torrance, CA, USA). The mobile phase was composed of 0.1% formic acid in water (solvent A) and methanol (solvent B). The elution gradient was as follows: from 5%B to 25%B for 6 min, 25%B to 5%B for 1 min, and held at 5%B for a further 3 min. A flow rate of 0.5 mL/min was used throughout. The column compartment and sample tray were set to 50 °C and 4 °C, respectively. The HPLC system was coupled to an Agilent Technologies 6210 LC/TOF mass spectrometer with a 1:4 post-column split. Mass spectrometric detection was conducted using APCI in the positive ion mode. MS experiments were carried out in the full scan mode, at 0.86 spectra/second, for the detection of $[M–H_2O+H]^+$ ions. Drying and nebulizing gases were set to 10 L/min and 25 psi, respectively, and a drying gas temperature of 330 °C was used throughout. The vaporizer and corona were set to 350 °C and 4 μA respectively, and a capillary voltage of 3,500 V was also used. Fragmentor and OCT 1 RF voltages were each set to 135 V, while the skimmer voltage was set to 50 V. Data acquisition and processing were performed by the MassHunter™ software package (Agilent™ Technologies Inc., Santa Clara, CA, USA). Metabolites were quantified via 10-point calibration curves of authentic standard compounds for which the R^2 coefficients were ≥ 0.99. The *p*-coumaraldehyde content was estimated by integrating the area of the mass peak eluting at Rt = 8.6 min ($[M-H]^-$ = 131.050238) and for which the ratio [theoretical mass/observed mass] was less than ±5 ppm (Fig. 26).

Carbohydrate and lignin assays

[0002] For each genotype (wild type, *C4H::qsuB-1,* and *C4H::qsuB*-9), samples consisted of equal mixtures of stem material from three independent cultures. Biomass was extracted

Date Reçue/Date Received 2020-05-29

sequentially by sonication (20 min) with 80% ethanol (three times), acetone (one time), chloroform-methanol (1:1, v/v, one time) and acetone (one time). For determination of carbohydrate composition, the biomass was acid-hydrolyzed as previously described (44). After $CaCO_3$ neutralization, monomeric sugars from the biomass hydrolyzates were separated by high-performance anion exchange chromatography with pulsed amperiometric detection using a PA20 column (Dionex, Sunnyvale, CA, USA) and quantified as previously described (45). A calibration curve of monosaccharide standards was run for verification of response factors. The standard NREL biomass protocol was used to measure lignin and ash (46). All carbohydrate and lignin assays were conducted in triplicate. The thioacidolysis procedure was carried out as described (47, 48) and the lignin-derived monomers were identified by GC-MS as their trimethyl-silylated derivatives.

2D ^{13}C-^{1}H heteronuclear single quantum coherence (HSQC) NMR spectroscopy

[0175] For each genotype (wild type, *C4H::qsuB-1* and *C4H::qsuB*-9), samples consisted of equal mixtures of stem material from three independent cultures. Samples were extracted and ball milled as previously described (49, 50). The gels were formed using DMSO-d_6/pyridine-d_5 (4:1) and sonicated until homogenous in a Branson™ 2510 table-top cleaner (Branson Ultrasonic Corporation, Danbury, CT). The temperature of the bath was closely monitored and maintained below 55 °C. The homogeneous solutions were transferred to NMR tubes. HSQC spectra were acquired at 25 °C using a Bruker™ Avance-600 MHz instrument equipped with a 5 mm inverse-gradient $^{1}H/^{13}C$ cryoprobe using a hsqcetgpsisp2.2 pulse program (ns = 400, ds = 16, number of increments = 256, d_1 = 1.0 s) (53). Chemical shifts were referenced to the central DMSO peak (δ_C/δ_H 39.5/2.5 ppm). Assignment of the HSQC spectra was described elsewhere (51, 54). A semi-quantitative analysis of the volume integrals of the HSQC correlation peaks was performed using Bruker's Topspin 3.1 (Windows) processing software. A Guassian apodization in F_2 (LB = -0.50, GB = 0.001) and squared cosine-bell in F_1 (LB = -0.10, GB = 0.001) were applied prior to 2D Fourier Transformation.

Isolation of cellulolytic enzyme lignin

[0176] For each genotype (wild type, *C4H::qsuB-1* and *C4H::qsuB*-9), samples consisted of equal mixtures of stem material from three independent cultures. The extracted biomass was ball-milled for 3 h per 500 mg of sample (in 10 min on/10 min off cycles) using a PM100 ball mill (Retsch, Newtown, PA) vibrating at 600 rpm in zirconium dioxide vessels (50 mL) containing ZrO_2 ball bearings (10 × 10 mm). Ball-milled walls were digested four

Date Reçue/Date Received 2020-05-29

times over 3 d at 50 °C with the polysaccharidases Cellic *CTec2* and HTec2 (Novozymes, Davis, CA) and pectinase from *Aspergillus niger* (Sigma-Aldrich, St. Louis, MO) in sodium citrate buffer (pH 5.0). The obtained cellulolytic lignin was washed with deionized water and lyophilized overnight.

Size exclusion chromatography

[0177] Lignin solutions, 1% (w/v), were prepared in analytical-grade 1-methyl-2-pyrrolidinone (NMP). The polydispersity of dissolved lignin was determined using analytical techniques involving SEC UV-$F_{250/400}$ as previously described (53). An Agilent 1200 series binary LC system (G1312B) equipped with diode-array (G1315D) and fluorescence (G1321A) detectors was used. Separation was achieved with a Mixed-D column (5 µm particle size, 300 mm x 7.5 mm i.d., linear molecular mass range of 200 to 400,000 u, Agilent Technologies Inc.) at 80 °C using a mobile phase of NMP at a flow rate of 0.5 ml/min. Absorbance of materials eluting from the column was detected using UV-F fluorescence (Ex_{250}/Em_{450}). Spectral intensities were area-normalized and molecular mass estimates were determined after calibration of the system with polystyrene standards.

Cell wall pretreatments and saccharification

[0178] Ball-milled senesced stems (10 mg) were mixed with 340 µL of water, 340 µL of H_2SO_4 (1.2%, w/v), or 340 µL of NaOH (0.25%, w/v) for hot water, dilute acid, or dilute alkali pretreatments, respectively; shaken at 1,400 rpm (30 °C, 30 min), and autoclaved at 120 °C for 1 h. Samples pretreated with dilute acid were neutralized with 5 N NaOH (25 µL). Saccharification was initiated by adding 650 µL of 100 mM sodium citrate buffer pH 5 (for hot water- and dilute alkali-pretreated samples) or 625 µL of 80 mM sodium citrate buffer pH 6.2 (for dilute acid-pretreated samples) containing 80 µg/mL tetracycline and 1% w/w or 0.2% w/w Cellic *CTec2* cellulase (Novozymes, Davis, CA). After 72 h of incubation at 50 °C with shaking (800 rpm), samples were centrifuged (20,000 × *g*, 3 min) and 10 µL of the supernatant was collected for measurement of reducing sugars using the 3,5-dinitrosalicylic acid assay and glucose solutions as standards (54).

Subcellular localization of QsuB

[0179] The *schl-qsuB* nucleotide sequence from the pTkan-*pC4H::schl-qsuB* construct was amplified using oligonucleotides 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTTCGATCTCCTCCT-3' (attB1 site underlined) and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCGTTTGGGATACCTCTCTCTAAATCT

C-3' (attB2 site underlined) and cloned into the Gateway pDONR221-f1 entry vector (Lalonde S, et al. (2010) *Front Physiol* 1:24). A sequence-verified entry clone was LR recombined with the pTKan-*p35S*-GWR1R2-YFP vector to generate the pTKan-*p35S-schl-qsuB-YFP* construct. Infiltration of 4-wo *N. benthamiana* leaves was done using the *Agrobacterium* strain GV3101, following the method described by Sparkes *et al.* (*Nat Protoc* 1(4):2019–2025). Plants transiently expressing the SCHL-QsuB-YFP fusion protein were analyzed by confocal laser scanning microscopy 2 d after the infiltration. The microscopy was performed using a Zeiss LSM 710 device (Carl Zeiss Microscopy, Jena, Germany) equipped with an argon laser (excitation at 514 nm and emission collected at 510 to 545 nm).

Lignin histochemical staining

[0180] Histochemical staining was performed as described by Pradhan-Mitra and Loqué ("Histochemical staining of *Arabidopsis thaliana* secondary cell wall elements," *JoVE* (*in press*)). Basal stem transverse sections (100 μm thick) were obtained using a vibratome. Sections were incubated for 3 min in phloroglucinol-HCl reagent (VWR International, Brisbane, CA), rinsed with water, and observed using bright field light microscopy (Leica Microsystems Inc., Buffalo Grove, IL).

Pyrolysis-gas chromatography mass spectrometry

[0181] Chemical composition of lignin in plant cell-wall samples were analyzed by pyrolysis-gas chromatography (GC)/mass spectrometry (MS) using a previously described method with some modifications (Del Río JC, *et al.* (2012) *J AgricFood Chem* 60(23):5922–5935). Pyrolysis of plant cell walls was performed with a Pyroprobe 5200 (CDS Analytical, Inc.) connected with GC/MS (Thermo Electron Corporation with Trace GC Ultra and Polaris-Q MS) equipped with an Agilent HP-5MS column (30 m × 0.25 mm i.d., 0.25 μm film thickness). The pyrolysis was carried out at 550 °C. The chromatograph was programmed from 50 °C (1 min) to 300 °C at a rate of 30 °C/min; the final temperature was held for 10 min. Helium was used as the carrier gas at a constant flow rate of 1 mL/min. The mass spectrometer was operated in scan mode and the ion source was maintained at 300 °C. The compounds were identified by comparing their mass spectra with those of the NIST library and those previously reported (Del Río JC, Gutiérrez A. (2006) *J Agric Food Chem* 54(13):4600–4610; Ralph J, Hatfield RD (1991) *J Agric Food Chem* 39(8):1426-1437). Peak molar areas were calculated for the lignin degradation products, the summed areas were normalized. Analyses on all samples were conducted in duplicate and data were averaged and expressed as percentages.

In vivo HCT activity assay

[0182] For the cloning of AtHCT, total *Arabidopsis* RNA (1 μg) were extracted using the Plant RNeasy extraction kit (Qiagen, Valencia, CA) and reverse-transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science, Indianapolis, IN). The obtained cDNA preparation was used to amplify AtHCT (GenBank accession number *NP_199704.1*) using the following oligonucleotides 5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT C ATGAAAATTA ACATCAGAGA TTCC-3' (attB1 site underlined) and 5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTCTCATATCTCAAACAAAAACTTCTCAAAC-3' (attB2 site underlined) for cloning into the Gateway pDONR221-f1 entry vector by BP recombination (Life Technologies, Foster City, CA). A sequence-verified *AtHCT* entry clone was LR recombined with the pDRf1-*4CL5*-GW vector (41) to generate the pDRf1-*4CL5-AtHCT* construct.

[0183] For For HCT activity assays, the pDRf1-*4CL5-AtHCT* and pDRf1-*4CL5* vectors were transformed into the *S. cerevisiae pad1* knockout (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 Δpad1*, ATCC 4005833) as previously described (41). Overnight cultures from single colonies harboring the pDRf1-*4CL5-AtHCT* and pDRf1-*4CL5* vectors were grown in 2X yeast nitrogen base medium without amino acids (Difco, Detroit, MI) supplemented with 6% glucose and 2X dropout mix without uracil (Sunrise Science Products, San Diego, CA). Overnight cultures were used to inoculated 10 mL of fresh minimal medium at an OD_{600} = 0.1. Substrates (*p*-coumarate, catechol or benzoates) were added to the medium 4 h later at a final concentration of 1 mM and the cultures were grown for 22 h. For the detection of the coumarate conjugate products, an aliquot of the culture medium was collected, cleared by centrifugation (20,000 × *g* for 5 min at 4 °C), mixed with an equal volume of 50% (v/v) methanol water and filtered using Amicon Ultra centrifugal filters (3,000 Da MW cutoff regenerated cellulose membrane; Millipore, Billerica, MA) prior to HPLC-ESI-TOF MS analysis.

REFERENCES

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

2. Boudet A-M (2007) Evolution and current status of research in phenolic compounds. *Phytochemistry* 68(22-24):2722–2735.

3. Keasling JD (2010) Manufacturing molecules through metabolic engineering. *Science* 330(6009):1355–1358.

4. Baucher M, Halpin, C, Petit-Conil, M, Boerjan W (2003) Lignin: Genetic engineering and impact on pulping. *Crit Rev Biochem Mol Biol* 38(4):305–350.

5. Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* 25(7):759–761.

6. Taboada A, et al. (2010) Digestibility of silages in relation to their hydroxycinnamic acid content and lignin composition. *J Sci Food Agric* 90(7):1155–1162.

7. Fraser CM, Chapple C (2011) The phenylpropanoid pathway in Arabidopsis. *The Arabidopsis Book* 9:e152.

8. Tohge T, Watanabe M, Hoefgen R, Fernie AR (2013) Shikimate and phenylalanine biosynthesis in the green lineage. Front Plant Sci 4:62.

9. Umezawa T (2010) The cinnamate/monolignol pathway. *Phytochemistry Rev* 9(1):1–17.

10. Vanholme R, et al. (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. Science 341(6150):1103–1106.

11. Li X, Weng J-K., Chapple C (2008) Improvement of biomass through lignin modification. *Plant J* 54(4):569–581.

12. Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Curr Opin Plant Biol 11(3):278–285.

13. Bonawitz ND, Chapple C. (2013) Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty? Curr Opin Biotechnol 24(2):336–343.

14. Voelker SL, Lachenbruch B, Meinzer FC, Kitin P, Strauss SH (2011) Transgenic poplars with reduced lignin show impaired xylem conductivity, growth efficiency and survival. *Plant Cell Environ* 34(4):655–668.

15. Brosnan CA, Voinnet O (2011) Cell-to-cell and long-distance siRNA movement in plants: mechanisms and biological implications. *Curr Opin Plant Biol* 14(5):580–587.

16. Iwase A, Matsui K, Ohme-Takagi M (2009) Manipulation of plant metabolic pathways by transcription factors. *Plant Biotechnol* 26(1):29–38.

17. Fornalé S, et al. (2010) ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. Plant J 64(4):633–644.

18. Shen H, et al. (2012) Functional characterization of the switchgrass (Panicum virgatum) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. New Phytol 193(1):121–136.

19. Yan L, et al. (2013) The heterologous expression in Arabidopsis thaliana of sorghum transcription factor SbbHLH1 downregulates lignin synthesis. *J Exp Bot* 64(10):3021–3302.

20. Costa MA, et al. (2013) Transgenic Hybrid Poplar for Sustainable and Scalable Production of the Commodity/Specialty Chemical, 2-Phenylethanol. *PloS ONE* 8(12):e83169.

21. Eudes A, et al. (2012) Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. Plant Biotechnol J 10(5):609–620.

22. Koeduka T, et al. (2013) Enhancement of production of eugenol and its glycosides in transgenic aspen plants via genetic engineering. *Biochem Biophys Res Commun* 436(1):73–78.

23. Zhang K, et al. (2012) An engineered monolignol 4-o-methyltransferase depresses lignin biosynthesis and confers novel metabolic capability in Arabidopsis. Plant Cell 24(7):3135–3152.

24. Zhang X, Gou M, Liu CJ (2014) Arabidopsis kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. Plant Cell 25(12):4994–*5010*.

25. Teramoto H, Inui M, Yukawa H (2009) Regulation of expression of genes involved in quinate and shikimate utilization in Corynebacterium glutamicum. *Appl Environ Microbiol* 75(11):3461–3468.

26. Eudes A, Liang Y, Mitra P, Loqué D. (2014) Lignin bioengineering. Curr Opin Biotechnol 16 *(in press)*.

27. Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW (2012) The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng* 109(4):1083–1087.

28. Yang F, et al (2013) Engineering secondary cell wall deposition in plants. Plant Biotechnol J 11(3):325–335.

29. Lin HH, Chen JH, Huang CC, Wang CJ (2007) Apoptotic effect of 3,4-dihydroxybenzoic acid on human gastric carcinoma cells involving JNK/p38 MAPK signaling activation. Int J Cancer 120(11):2306–2316.

30. Otsuka Y, et al. (2006) Efficient production of 2-pyrone 4,6-dicarboxylic acid as a novel polymer-based material from protocatechuate by microbial function. Appl Microbiol Biotechnol 71(5):608–614.

31. Sangha AK, et al. (2014) Chemical Factors that Control Lignin Polymerization. J Phys Chem B 118(1):164–170.

32. Ziebell A, et al. (2010) Increase in 4-coumaryl alcohol units during lignification in alfalfa (Medicago sativa) alters the extractability and molecular weight of lignin. J Biol Chem 285(50):38961–38968.

33. Ralph J, et al. (2006) Effects of coumarate 3-hydroxylase down-regulation on lignin structure. J Biol Chem 281(13):8843–8853.

34. Moglia A, et al (2010) Production of novel antioxidative phenolic amides through heterologous expression of the plant's chlorogenic acid biosynthesis genes in yeast. Metab Eng 12(3):223–232.

35. Sander M, Petersen M (2011) Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl-transferase, from Coleus blumei Benth. *Planta* 233(6):1157–1171.

36. Yuan L, et al. (2009) AtAMT1;4, a pollen-specific high-affinity ammonium transporter of the plasma membrane in Arabidopsis. Plant Cell Physiol 50(1):13–25.

37. Kim JG et al. (2009) Xanthomonas T3S Effector XopN Suppresses PAMP-Triggered Immunity and Interacts with a Tomato Atypical Receptor-Like Kinase and TFT1. Plant Cell 21(4):1305–1323.

38. Lebrun M, Leroux B, Sailland A (1992) Gène chimère pour la transformation des plantes. European patent application. Patent Application No. EP 508909A1.

39. Bechtold N, Pelletier G (1998) In planta Agrobacterium-mediated transformation of adult Arabidopsis thaliana plants by vacuum infiltration. *Methods in molecular biology (Clifton, NJ)* 82:259–266.

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

41. Eudes A, et al. (2011) Production of tranilast [N-(3',4'-dimethoxycinnamoyl)-anthranilic acid] and its analogs in yeast Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* 89(4):989–1000.

42. Bokinsky G, et al. (2013) HipA-triggered growth arrest and β-lactam tolerance in Escherichia coli are mediated by RelA-dependent ppGpp synthesis. J Bacteriol 195(14):3173–3182.

43. Eudes A, et al. (2013) Production of hydroxycinnamoyl anthranilates from glucose in Escherichia coli. *Microb Cell Fact* 12:62.

44. Moxley G, Zhang YHP (2007) More accurate determination of acid-labile carbohydrate composition in lignocellulose by modified quantitative saccharification. *Energy Fuels* 21(6):3684–3688.

45. ØBro J, Harholt J, Scheller H, Orfila (2004) Rhamnogalacturonan I in Solanum tuberosum tubers contains complex arabinogalactan structures. *Phytochemistry* 65(10):1429–1438.

46. Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J (2008) Determination of structural carbohydrates and lignin in biomass. In *Laboratory Analytical Procedure* (Technical Report, NREL/TP-510-42618), Golden, CO: National Renewable Energy Laboratory.

47. Lapierre C, Pollet B, Rolando C (1995) New insights into the molecular architecture of hardwood lignins by chemical degradative methods. *Res Chem Intermed* 21(3-5):397–412.

48. Lapierre C, et al. (1999) Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiol* 119(1):153–164.

49. Kim H, Ralph J (2010) Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d(6)/pyridine-d(5). *Org Biomol Chem* 8(3):576–591.

50. Mansfield SD, Kim H, Lu F, Ralph J (2012) Whole plant cell wall characterization using solution-state 2D NMR. *Nat Protoc* 7(9):1579–1589.

51. Heikkinen S, Toikka MM, Karhunen PT, Kilpeläinen IA (2003) Quantitative 2D HSQC (Q-HSQC) via suppression of J-dependence of polarization transfer in NMR spectroscopy: application to wood lignin. *J Am Chem Soc* 125(14):4362–4367.

52. Yelle DJ, Ralph J, Frihart CR (2008) Characterization of nonderivatized plant cell walls using high-resolution solution-state NMR spectroscopy. *Magn Reson Chem* 46(6):508–517.

53. George A, et al. (2011) The effect of ionic liquid cation and anion combinations on the macromolecular structure of lignins. *Green Chem* 13:3375–*3385*.

54. Miller G (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* 31(3):426–428.

ILLUSTRATIVE SEQUENCES

SEQ ID NO:1 – MtAroK polynucleotide sequence

ATGGCACCAAAAGCTGTTTTAGTGGGACTTCCTGGAAGTGGAAAGTCCACTATCGGTAGAAG
GTTGGCTAAAGCATTAGGAGTTGGTTTGTTAGACACTGATGTGGCTATAGAACAAAGGACAG
GAAGATCAATAGCAGACATTTTTGCTACAGATGGTGAACAGGAGTTCAGAAGGATAGAAGAG
GATGTTGTGAGAGCTGCATTGGCTGACCATGATGGTGTTCTTAGTTTGGGTGGAGGTGCAGT
TACTTCCCCAGGAGTGAGAGCTGCACTTGCTGGTCACACAGTTGTGTATTTGGAAATCTCAG
CTGCAGAGGGAGTGAGAAGGACAGGTGGTAACACCGTGAGACCACTTTTGGCAGGTCCTGAT

AGGGCTGAAAAGTATAGAGCTTTGATGGCAAAAAGGGCTCCTTTATACAGAAGGGTTGCTAC
TATGAGAGTGGATACAAATAGAAGGAACCCAGGTGCAGTTGTTAGGCACATTTTATCCAGGT
TGCAGGTTCCATCTCCTTCTGAGGCAGCTACT

SEQ ID NO:2 – MtAroK amino acid sequence (*Mycobacterium tuberculosis* shikimate kinase; NP_217055)
MAPKAVLVGLPGSGKSTIGRRLAKALGVGLLDTDVAIEQRTGRSIADIFATDGEQEFRRIEE
DVVRAALADHDGVLSLGGGAVTSPGVRAALAGHTVVYLEISAAEGVRRTGGNTVRPLLAGPD
RAEKYRALMAKRAPLYRRVATMRVDTNRRNPGAVVRHILSRLQVPSPSEAAT

SEQ ID NO:3 – ScAro1 polynucleotide sequence
ATGGTTCAGCTTGCTAAGGTGCCTATTTTGGGTAACGACATCATTCACGTTGGATATAACAT
TCACGATCATTTGGTTGAGACTATTATCAAGCATTGTCCATCTTCTACTTATGTTATTTGTA
ACGATACCAACCTTTCTAAGGTTCCTTATTACCAACAGTTAGTGCTTGAGTTTAAGGCTTCT
TTGCCAGAAGGAAGTAGATTGTTAACTTATGTTGTGAAACCTGGAGAGACTTCTAAGTCAAG
GGAAACAAAAGCTCAATTGGAGGACTACCTTTTGGTTGAAGGATGTACCAGAGATACTGTGA
TGGTTGCTATTGGTGGAGGTGTTATAGGTGATATGATTGGATTTGTGGCATCAACTTTCATG
AGAGGTGTTAGGGTTGTGCAAGTGCCAACAAGTTTACTTGCTATGGTTGACAGTTCCATCGG
AGGAAAGACAGCAATAGATACCCCATTGGGAAAAAACTTTATTGGTGCTTTCTGGCAGCCTA
AGTTCGTGCTTGTTGATATCAAGTGGCTTGAGACATTGGCTAAGAGAGAATTTATCAACGGA
ATGGCAGAAGTTATCAAGACAGCTTGTATTTGGAACGCAGATGAGTTTACCAGATTGGAATC
AAATGCTAGTTTGTTCTTAAACGTTGTGAACGGTGCAAAGAACGTGAAGGTTACTAACCAAC
TTACAAACGAGATCGATGAAATCTCAAATACCGACATCGAAGCTATGCTTGATCACACTTAC
AAACTTGTTTTGGAGTCTATCAAGGTGAAAGCAGAAGTTGTGTCTTCAGATGAGAGAGAAAG
TTCCTTGAGGAACTTGCTTAACTTCGGTCATTCAATCGGACACGCTTACGAAGCAATCTTAA
CTCCACAAGCTCTTCATGGAGAATGTGTTTCTATTGGTATGGTGAAGGAGGCAGAATTGTCA
AGATACTTCGGAATATTAAGTCCTACACAGGTTGCAAGGTTGTCCAAAATTTTGGTTGCTTA
CGGTTTGCCAGTGTCTCCTGATGAGAAGTGGTTCAAGGAATTAACACTTCATAAAAAGACCC
CTTTAGACATCCTTTTGAAAAAGATGTCCATCGATAAAAAGAATGAGGGTTCTAAAAAGAAA
GTTGTGATCTTAGAATCTATCGGAAAGTGCTATGGAGACTCCGCTCAATTTGTTTCTGATGA
GGACCTTAGATTCATTTTGACAGATGAAACCCTTGTTTACCCATTTAAAGATATACCTGCTG
ACCAACAGAAGGTTGTGATTCCACCTGGTAGTAAATCCATTTCTAACAGAGCATTGATCTTA
GCTGCATTGGGTGAAGGACAGTGTAAGATAAAGAACCTTCTTCATTCAGATGACACTAAGCA
CATGCTTACAGCAGTTCATGAATTGAAAGGTGCTACAATCTCTTGGGAGGATAACGGAGAAA
CCGTTGTGGTTGAAGGTCATGGAGGTTCCACTTTGTCTGCTTGCGCAGATCCACTTTATTTG
GGTAATGCTGGAACCGCATCAAGATTTTTAACTAGTCTTGCTGCTTTGGTTAACTCAACTTC
TTCACAAAAGTACATTGTGTTAACTGGTAATGCAAGAATGCAACAGAGGCCAATCGCTCCTT
TAGTTGATTCTCTTAGAGCAAACGGAACAAAGATCGAGTACCTTAACAACGAAGGTTCACTT
CCTATCAAGGTTTACACTGATAGTGTGTTCAAAGGAGGTAGAATAGAATTAGCTGCAACAGT
TAGTTCCCAATATGTGTCTTCAATTCTTATGTGTGCTCCATACGCAGAAGAGCCTGTTACTT
TAGCTCTTGTGGGAGGAAAGCCAATCTCAAAATTGTACGTTGATATGACAATCAAGATGATG
GAAAAGTTCGGAATCAACGTTGAGACTTCTACTACAGAACCATACACATACTACATCCCTAA
GGGTCATTACATCAACCCTTCAGAGTACGTTATCGAAAGTGATGCTAGTTCCGCAACTTATC
CATTAGCTTTCGCTGCAATGACCGGAACCACTGTGACTGTTCCTAATATTGGATTTGAATCT
CTTCAAGGTGACGCTAGATTCGCAAGGGATGTTTTGAAGCCAATGGGTTGTAAAATCACTCA
GACAGCTACCTCAACAACCGTTAGTGGTCCACCTGTGGGAACATTAAAGCCACTTAAACACG
TTGACATGGAACCTATGACAGATGCTTTCTTGACCGCATGTGTGGTTGCTGCAATTTCACAT
GATAGTGACCCAAATTCTGCTAACACTACAACCATAGAGGGAATAGCAAACCAAAGAGTTAA
GGAATGCAACAGGATCTTGGCTATGGCAACTGAGTTAGCTAAATTTGGTGTTAAAACTACAG
AATTACCTGATGGAATCCAGGTGCACGGTCTTAATTCAATCAAGGACTTGAAAGTTCCAAGT
GATTCTTCAGGTCCTGTGGGAGTTTGTACTTATGATGACCATAGAGTGGCAATGTCATTCAG
TTTGTTAGCTGGTATGGTTAATTCTCAAAACGAGAGGGATGAAGTGGCTAACCCAGTTAGAA
TTTTGGAAAGGCACTGCACTGGAAAGACATGGCCTGGTTGGTGGACGTTTTGCATAGTGAA

TTAGGAGCTAAACTTGATGGTGCAGAGCCTTTAGAATGTACTTCTAAGAAAAATTCCAAGAA
ATCTGTGGTTATTATCGGAATGAGAGCTGCAGGTAAAACCACTATTTCCAAATGGTGCGCTT
CTGCATTGGGATACAAATTGGTTGATTTAGACGAGCTTTTTGAACAACAGCATAATAACCAA
TCAGTTAAGCAGTTCGTGGTTGAGAACGGTTGGGAAAAATTTAGAGAAGAGGAAACTAGGAT
CTTCAAGGAAGTTATCCAAAACTACGGTGATGACGGATACGTTTTCTCTACAGGAGGTGGAA
TTGTGGAGTCAGCTGAAAGTAGAAAGGCACTTAAAGATTTCGCTAGTTCCGGTGGATATGTG
TTGCATTTACACAGGGACATTGAGGAAACTATCGTTTTCTTGCAATCTGATCCATCAAGACC
AGCTTATGTTGAGGAAATTAGAGAAGTGTGGAACAGAAGGGAGGGTTGGTACAAGGAATGTT
CAAACTTCTCTTTCTTTGCTCCACACTGCTCTGCTGAGGCAGAATTTCAAGCTCTTAGAAGG
TCCTTCTCTAAATACATCGCAACTATAACAGGAGTTAGAGAGATCGAAATACCATCCGGTAG
GTCTGCTTTTGTTTGTTTGACCTTCGATGACTTAACCGAGCAGACTGAAAACTTAACTCCTA
TTTGTTATGGTTGCGAGGCAGTGGAAGTTAGAGTGGACCATCTTGCTAATTACTCAGCAGAT
TTCGTTTCCAAGCAATTGTCTATCCTTAGAAAGGCTACTGATAGTATCCCAATAATTTTCAC
AGTTAGGACCATGAAACAGGGTGGAAACTTTCCTGACGAGGAATTTAAGACACTTAGAGAAT
TGTACGATATAGCTCTTAAGAATGGTGTTGAGTTTCTTGACTTGGAATTAACTCTTCCTACA
GATATCCAATACGAAGTTATCAACAAGAGAGGAAACACTAAGATCATAGGTTCCCATCACGA
TTTTCAAGGATTATACTCTTGGGATGACGCTGAGTGGGAAAATAGATTCAACCAGGCATTGA
CCTTAGATGTTGACGTGGTTAAGTTTGTGGGTACTGCTGTTAATTTCGAGGACAACCTTAGA
TTGGAACATTTTAGGGATACACACAAGAACAAGCCACTTATCGCAGTTAACATGACCTCAAA
AGGATCAATCAGTAGAGTGTTGAATAACGTTTTAACCCCTGTGACTTCCGATCTTTTGCCAA
ACTCTGCTGCACCTGGTCAACTTACCGTTGCTCAGATCAACAAGATGTACACTTCTATGGGT
GGAATTGAGCCAAAAGAACTTTTCGTGGTTGGAAAGCCAATCGGACATTCAAGATCACCTAT
CTTGCATAACACTGGATACGAAATTTTAGGTCTTCCTCATAAGTTCGATAAATTCGAGACAG
AATCTGCTCAATTGGTTAAGGAAAAATTACTTGATGGTAACAAGAACTTTGGTGGAGCTGCA
GTTACTATCCCATTGAAATTGGATATCATGCAGTACATGGATGAATTGACAGACGCTGCAAA
GGTTATTGGTGCTGTGAATACCGTTATCCCACTTGGAAACAAGAAGTTCAAGGGTGATAACA
CAGACTGGCTTGGAATAAGAAATGCTCTTATCAACAACGGTGTTCCTGAATATGTGGGTCAC
ACTGCAGGATTGGTTATTGGTGCTGGTGGAACATCAAGAGCTGCATTATACGCTCTTCATAG
TTTGGGTTGTAAGAAAATCTTTATCATCAACAGGACAACCTCTAAGTTAAAACCACTTATCG
AGTCACTTCCTAGTGAATTTAACATCATCGGAATAGAGTCCACTAAGTCTATTGAGGAAATC
AAAGAACACGTTGGTGTGGCAGTTTCCTGCGTTCCAGCTGATAAACCTTTGGATGACGAGTT
GCTTTCAAAACTTGAAAGATTTTTGGTTAAGGGTGCTCATGCTGCATTCGTGCCAACACTTT
TGGAAGCTGCATATAAGCCATCCGTGACCCCTGTTATGACTATCTCTCAGGATAAGTACCAG
TGGCACGTGGTTCCTGGATCTCAAATGTTGGTTCATCAGGGTGTGGCTCAGTTTGAGAAGTG
GACAGGATTCAAAGGACCATTTAAGGCTATTTTCGACGCAGTTACCAAGGAG

SEQ ID NO:4 – ScAro1 amino acid sequence (*Saccharomyces cerevisiae* Pentafunctional arom protein; CAA88208)

MVQLAKVPILGNDIIHVGYNIHDHLVETIIKHCPSSTYVICNDTNLSKVPYYQQLVLEFKAS
LPEGSRLLTYVVKPGETSKSRETKAQLEDYLLVEGCTRDTVMVAIGGGVIGDMIGFVASTFM
RGVRVVQVPTSLLAMVDSSIGGKTAIDTPLGKNFIGAFWQPKFVLVDIKWLETLAKREFING
MAEVIKTACIWNADEFTRLESNASLFLNVVNGAKNVKVTNQLTNEIDEISNTDIEAMLDHTY
KLVLESIKVKAEVVSSDERESSLRNLLNFGHSIGHAYEAILTPQALHGECVSIGMVKEAELS
RYFGILSPTQVARLSKILVAYGLPVSPDEKWFKELTLHKKTPLDILLKKMSIDKKNEGSKKK
VVILESIGKCYGDSAQFVSDEDLRFILTDETLVYPFKDIPADQQKVVIPPGSKSISNRALIL
AALGEGQCKIKNLLHSDDTKHMLTAVHELKGATISWEDNGETVVVEGHGGSTLSACADPLYL
GNAGTASRFLTSLAALVNSTSSQKYIVLTGNARMQQRPIAPLVDSLRANGTKIEYLNNEGSL
PIKVYTDSVFKGGRIELAATVSSQYVSSILMCAPYAEEPVTLALVGGKPISKLYVDMTIKMM
EKFGINVETSTTEPYTYYIPKGHYINPSEYVIESDASSATYPLAFAAMTGTTVTVPNIGFES
LQGDARFARDVLKPMGCKITQTATSTTVSGPPVGTLKPLKHVDMEPMTDAFLTACVVAAISH
DSDPNSANTTTIEGIANQRVKECNRILAMATELAKFGVKTTELPDGIQVHGLNSIKDLKVPS
DSSGPVGVCTYDDHRVAMSFSLLAGMVNSQNERDEVANPVRILERHCTGKTWPGWWDVLHSE
LGAKLDGAEPLECTSKKNSKKSVVIIGMRAAGKTTISKWCASALGYKLVDLDELFEQQHNNQ

SVKQFVVENGWEKFREEETRIFKEVIQNYGDDGYVFSTGGGIVESAESRKALKDFASSGGYV
LHLHRDIEETIVFLQSDPSRPAYVEEIREVWNRREGWYKECSNFSFFAPHCSAEAEFQALRR
SFSKYIATITGVREIEIPSGRSAFVCLTFDDLTEQTENLTPICYGCEAVEVRVDHLANYSAD
FVSKQLSILRKATDSIPIIFTVRTMKQGGNFPDEEFKTLRELYDIALKNGVEFLDLELTLPT
DIQYEVINKRGNTKIIGSHHDFQGLYSWDDAEWENRFNQALTLDVDVVKFVGTAVNFEDNLR
LEHFRDTHKNKPLIAVNMTSKGSISRVLNNVLTPVTSDLLPNSAAPGQLTVAQINKMYTSMG
GIEPKELFVVGKPIGHSRSPILHNTGYEILGLPHKFDKFETESAQLVKEKLLDGNKNFGGAA
VTIPLKLDIMQYMDELTDAAKVIGAVNTVIPLGNKKFKGDNTDWLGIRNALINNGVPEYVGH
TAGLVIGAGGTSRAALYALHSLGCKKIFIINRTTSKLKPLIESLPSEFNIIGIESTKSIEEI
KEHVGVAVSCVPADKPLDDELLSKLERFLVKGAHAAFVPTLLEAAYKPSVTPVMTISQDKYQ
WHVVPGSQMLVHQGVAQFEKWTGFKGPFKAIFDAVTKE

SEQ ID NO:5 -- CgQsuB polynucleotide sequence
ATGAGAACAAGTATTGCAACCGTTTGTTTATCCGGAACTCTTGCTGAAAAATTGAGAGCAGC
TGCAGACGCAGGATTCGATGGTGTTGAGATTTTTGAACAAGATTTGGTTGTGTCTCCACATT
CAGCTGAACAAATCAGACAGAGGGCACAAGATTTAGGTCTTACATTGGACTTATTTCAGCCT
TTCAGAGATTTTGAAGGAGTTGAAGAGGAACAATTCTTAAAGAATCTTCACAGGTTGGAGGA
AAAATTTAAGTTAATGAACAGACTTGGTATCGAAATGATCTTGCTTTGTTCTAACGTTGGAA
CAGCTACCATCAACGATGACGATCTTTTTGTGGAACAATTGCATAGAGCTGCAGATTTGGCT
GAGAAGTACAACGTTAAGATCGCTTATGAAGCTCTTGCTTGGGGTAAATTCGTTAATGATTT
TGAGCATGCTCACGCATTGGTTGAAAAAGTGAACCATAAGGCTTTGGGTACTTGCTTAGATA
CATTCCACATATTAAGTAGAGGATGGGAGACTGATGAGGTTGAAAACATCCCAGCTGAAAAA
ATATTTTTCGTGCAATTGGCTGATGCACCTAAGTTATCTATGGATATCCTTTCTTGGTCAAG
GCATCACAGAGTTTTTCCAGGAGAGGGTGACTTCGATTTGGTTAAGTTCATGGTGCATCTTG
CTAAGACAGGATACGATGGTCCTATATCTTTGGAGATTTTCAACGACTCATTTAGGAAAGCT
GAAGTTGGAAGAACTGCAATTGATGGTTTAAGGTCTCTTAGATGGTTGGAGGACCAAACATG
GCATGCACTTAACGCTGAAGATAGGCCATCAGCACTTGAGTTGAGAGCTTTGCCAGAAGTTG
CAGAGCCTGAGGGTGTGGATTTCATTGAGATCGCTACAGGAAGGTTAGGTGAAACCATCAGA
GTTTTACACCAGCTTGGTTTTAGACTTGGTGGACATCACTGTTCTAAGCAGGATTATCAAGT
TTGGACTCAAGGAGATGTGAGGATCGTTGTGTGCGACAGAGGAGCAACAGGTGCTCCTACCA
CTATATCAGCTATGGGTTTCGACACCCCAGATCCTGAGGCTGCACATGCTAGGGCAGAACTT
TTGAGAGCACAAACAATTGATAGACCACACATCGAGGGAGAAGTTGATCTTAAGGGTGTGTA
CGCTCCTGACGGAGTTGAATTGTTTTTCGCAGGACCATCTCCTGATGGTATGCCAGAGTGGT
TACCTGAATTTGGTGTTGAGAAGCAAGAAGCTGGACTTATCGAAGCAATCGATCATGTTAAC
TTTGCTCAGCCTTGGCAACACTTCGATGAGGCAGTTTTGTTTTATACCGCATTGATGGCTTT
AGAAACTGTGAGAGAGGATGAATTTCCATCACCTATTGGTTTAGTTAGGAATCAGGTGATGA
GATCACCAAACGATGCTGTTAGATTACTTTTGTCAGTGGCACCTGAGGACGGAGAACAGGGT
GATTTCTTAAATGCTGCATACCCAGAACATATAGCTCTTGCAACTGCTGATATTGTTGCAGT
GGCTGAAAGAGCTAGGAAAAGAGGTTTGGATTTCTTGCCAGTTCCTGAAAACTATTACGACG
ATGTGCAGGCTAGATTCGATTTGCCTCAAGAGTTTTTAGACACACTTAAGGAAAACCATCTT
CTTTATGACTGCGATGAGAACGGTGAATTTTTGCACTTCTACACTAGAACATTGGGAACATT
ATTTTTCGAGGTTGTGGAAAGAAGGGGTGGATTTGCTGGATGGGGTGAAACCAATGCACCTG
TTAGGCTTGCTGCTCAATATAGAGAAGTTAGAGATTTAGAGAGAGGTATCCCAAAC

SEQ ID NO:6 – CgQsuB amino acid sequence (*Corynebacterium glutamicum* dehydroshikimate dehydratase; BAF53460)
MRTSIATVCLSGTLAEKLRAAADAGFDGVEIFEQDLVVSPHSAEQIRQRAQDLGLTLDLFQP
FRDFEGVEEEQFLKNLHRLEEKFKLMNRLGIEMILLCSNVGTATINDDDLFVEQLHRAADLA
EKYNVKIAYEALAWGKFVNDFEHAHALVEKVNHKALGTCLDTFHILSRGWETDEVENIPAEK
IFFVQLADAPKLSMDILSWSRHHRVFPGEGDFDLVKFMVHLAKTGYDGPISLEIFNDSFRKA
EVGRTAIDGLRSLRWLEDQTWHALNAEDRPSALELRALPEVAEPEGVDFIEIATGRLGETIR
VLHQLGFRLGGHHCSKQDYQVWTQGDVRIVVCDRGATGAPTTISAMGFDTPDPEAAHARAEL
LRAQTIDRPHIEGEVDLKGVYAPDGVELFFAGPSPDGMPEWLPEFGVEKQEAGLIEAIDHVN

FAQPWQHFDEAVLFYTALMALETVREDEFPSPIGLVRNQVMRSPNDAVRLLLSVAPEDGEQG
DFLNAAYPEHIALATADIVAVAERARKRGLDFLPVPENYYDDVQARFDLPQEFLDTLKENHL
LYDCDENGEFLHFYTRTLGTLFFEVVERRGGFAGWGETNAPVRLAAQYREVRDLERGIPN

SEQ ID NO:7 – PaDsDH polynucleotide sequence
ATGCCTTCAAAACTTGCTATCACCTCAATGTCTCTTGGTAGATGCTATGCTGGTCACTCCTT
CACTACTAAATTGGATATGGCTAGGAAATATGGTTACCAAGGACTTGAATTGTTCCATGAGG
ATTTGGCTGACGTTGCATATAGACTTAGTGGTGAAACACCATCCCCTTGTGGACCATCTCCT
GCTGCACAGTTGAGTGCTGCAAGACAAATACTTAGGATGTGTCAGGTTAGAAACATAGAAAT
TGTGTGCTTACAGCCATTTTCTCAATACGATGGTTTGTTAGACAGAGAAGAGCATGAAAGAA
GGCTTGAACAATTGGAGTTCTGGATAGAATTAGCTCACGAGCTTGATACAGACATTATCCAG
ATTCCAGCAAATTTTCTTCCTGCTGAAGAGGTTACCGAAGATATTTCTTTGATCGTTTCAGA
TTTGCAAGAGGTGGCTGACATGGGTTTGCAGGCAAACCCACCTATTAGATTCGTTTATGAAG
CTCTTTGTTGGTCAACTAGAGTGGATACATGGGAAAGGAGTTGGGAGGTTGTGCAAAGAGTT
AATAGGCCTAACTTTGGTGTGTGCCTTGATACATTCAATATCGCAGGAAGAGTTTACGCTGA
CCCAACCGTGGCATCAGGTAGAACTCCTAACGCTGAAGAGGCAATTAGGAAGTCAATCGCTA
GATTGGTTGAAAGGGTTGATGTTAGTAAAGTTTTCTATGTGCAAGTTGTGGACGCAGAGAAG
TTGAAAAAGCCATTAGTTCCTGGACACAGATTCTACGATCCAGAACAACCTGCTAGGATGTC
TTGGTCAAGAAACTGCAGGTTGTTTTATGGTGAAAAAGATAGAGGAGCTTACTTGCCAGTTA
AGGAGATTGCTTGGGCATTTTTCAATGGTTTGGGATTTGAAGGTTGGGTTTCCTTAGAGCTT
TTCAACAGAAGGATGTCTGATACTGGTTTTGGAGTGCCTGAAGAGTTAGCTAGAAGGGGAGC
AGTTTCCTGGGCTAAACTTGTGAGAGATATGAAGATCACCGTTGACTCACCAACTCAACAGC
AAGCTACACAGCAACCTATAAGAATGTTGAGTTTATCCGCTGCATTA

SEQ ID NO:8 – PaDsDH amino acid sequence (*Podospora anserina* dehydroshikimate dehydratase; CAD60599)
MPSKLAITSMSLGRCYAGHSFTTKLDMARKYGYQGLELFHEDLADVAYRLSGETPSPCGPSP
AAQLSAARQILRMCQVRNIEIVCLQPFSQYDGLLDREEHERRLEQLEFWIELAHELDTDIIQ
IPANFLPAEEVTEDISLIVSDLQEVADMGLQANPPIRFVYEALCWSTRVDTWERSWEVVQRV
NRPNFGVCLDTFNIAGRVYADPTVASGRTPNAEEAIRKSIARLVERVDVSKVFYVQVVDAEK
LKKPLVPGHRFYDPEQPARMSWSRNCRLFYGEKDRGAYLPVKEIAWAFFNGLGFEGWVSLEL
FNRRMSDTGFGVPEELARRGAVSWAKLVRDMKITVDSPTQQQATQQPIRMLSLSAAL

SEQ ID NO:9 – PhPAAS polynucleotide sequence
ATGGACACTATCAAGATCAACCCAGAGTTCGACGGACAGTTCTGCAAGACTACATCATTATT
AGACCCAGAGGAGTTCAGGAGGAATGGACATATGATGGTTGATTTTCTTGCTGACTACTTCC
ACAACATCGAAAAGTACCCAGTTAGATCCCAAGTGGAACCTGGTTATTTGGAGAGGTTGTTA
CCAGATTCAGCTCCTATACAGCCAGAACCTATCGAGAAAATTTTGAAGGATGTTAGATCAGA
CATATTTCCAGGTTTAACACATTGGCAAAGTCCAAATTTCTTTGCTTACTTCCCTTGCTCTT
CAAGTACCGCAGGAATTTTAGGTGAAATGCTTTCAGCTGGATTGAACGTTGTGGGTTTTTCA
TGGATCGCTAGTCCAGCTGCAACTGAATTAGAGAGTATTGTTATGGATTGGCTTGGAAAATT
GATTAATTTGCCTAAGACATATCTTTTCTCTGGTGGAGGTGGAGGTGTGATGCAGGGTACTA
CATGCCAAGTTATGCTTTGTACTATCGTGGCTGCAAGAGATAAAATGTTGGAAAAGTTTGGA
AGGGACAACATTGATAAGTTAGTTGTGTACGCATCAGACCAAACCCACTTTAGTTTCCAGAA
AGCTGTTAAGATCTCAGGTATAAAACCAGAAAACTTCAGAGCTATACCTACCACTAAGGCAA
CAGAATTCTCCCTTAACCCAGAGTCTTTGAGAAGGGCTATCCAAGAGGATAAAAAGGCAGGA
CTTATCCCTTTGTTTTTATGCACATCAATAGGTACAACCAGTACTACAGCAGTTGACCCACT
TAAACCTTTGTGTGAAATAGCTGAAGAGTATGGAATTTGGGTTCATGTGGATGCTGCATACG
CTGGTTCTGCATGCATTTGTCCTGAATTTCAGCATTTCTTGGACGGTGTTGAGCACGCTAAT
TCCTTTTCTTTCAACGCACACAAGTGGTTGTTTACTACTCTTGATTGTTGCTGTCTTTGGTT
GAAAGACCCATCCTCTTTGACTAAGGCACTTTCAACAAACCCTGAAGTTTTGAGAAACGATG
CTACCGACAGTGAGCAAGTTGTGGATTATAAAGACTGGCAGATTACTTTATCCAGAAGGTTT
AGGTCTCTTAAGCTTTGGTTGGTTCTTAAGTCCTACGGAGTGGCTAATCTTAGAAACTTCAT

AAGGTCTCATATCGAAATGGCTAAGCACTTTGAAGAGTTGGTTGCAATGGATGAAAGATTCG
AGATCATGGCACCAAGGAATTTTTCCTTAGTTTGTTTCAGAGTGTCTCTTTTGGCTCTTGAA
AAGAAGTTTAATTTCGTTGATGAAACTCAAGTGAACGAGTTTAACGCTAAGCTTCTTGAATC
TATCATCTCAAGTGGTAACGTTTACATGACACATACCGTTGTGGAGGGAGTTTACATGATTA
GATTCGCTGTGGGTGCACCTTTGACAGATTATCCTCACATTGATATGGCTTGGAATGTTGTT
AGGAACCACGCTACTATGATGTTGAACGCA

SEQ ID NO:10 – PhPAAS amino acid sequence (*Petunia hybrida* Phenylacetaldehyde synthase; ABB72475)

MDTIKINPEFDGQFCKTTSLLDPEEFRRNGHMMVDFLADYFHNIEKYPVRSQVEPGYLERLL
PDSAPIQPEPIEKILKDVRSDIFPGLTHWQSPNFFAYFPCSSSTAGILGEMLSAGLNVVGFS
WIASPAATELESIVMDWLGKLINLPKTYLFSGGGGGVMQGTTCEVMLCTIVAARDKMLEKFG
RENIDKLVVYASDQTHFSFQKAVKISGIKPENFRAIPTTKATEFSLNPESLRRAIQEDKKAG
LIPLFLCTSIGTTSTTAVDPLKPLCEIAEEYGIWVHVDAAYAGSACICPEFQHFLDGVEHAN
SFSFNAHKWLFTTLDCCCLWLKDPSSLTKALSTNPEVLRNDATDSEQVVDYKDWQITLSRRF
RSLKLWLVLKSYGVANLRNFIRSHIEMAKHFEELVAMDERFEIMAPRNFSLVCFRVSLLALE
KKFNFVDETQVNEFNAKLLESIISSGNVYMTHTVVEGVYMIRFAVGAPLTDYPHIDMAWNVV
RNHATMMLNA

SEQ ID NO:11 – ObCCMT1 polynucleotide sequence

ATGGCGAGAAAAGAGAACTATGTTGTTTCTAACATGAATGTTGAAAGTGTGTTGTGCATGAA
AGGTGGAAAAGGAGAAGATAGCTATGATAACAACTCTAAGATGCAGGAGCAACATGCTCGAT
CAGTGCTCCACCTTCTGATGGAAGCTCTCGACGGCGTGGGGCTGAGCTCGGTGGCGGCCGGC
GCTTTCGTGGTGGCGGATCTCGGCTGCTCCAGCGGAAGAAACGCCATAAACACGATGGAATT
TATGATCAATCACCTGACTGAGCACTACACGGTGGCGGCGGAAGAGCCGCCGGAATTCTCAG
CCTTCTTCTGCGACCTCCCCTCCAACGACTTCAACACCCTCTTTCAGCTCCTTCCGCCGTCT
GACGGCAGCAGCGGTTCTTACTTCACTGCCGGCGTGGCCGGTTCGTTTTACCGGAGGCTTTT
CCCGGCGAAGTCTGTTGATTTCTTTTACTCGGCATTTAGTTTGCACTGGCTATCTCAGATAC
CAAAGGAGGTGATGGAGAAGGGATCGGCGGCTTACAACGAGGGGAGAGTGACCATCAACGGT
GCAAAAGAGAGCACCGTAAATGCATACAAGAAACAATTTCAAAGTGATTTGGGTGTCTTCTT
GAGATCCAGATCCAAAGAATTGAAACCGGGAGGATCCATGTTCCTCATGCTCTTGGGTCGGA
CCAGCCCCGACCCGGCAGATCAGGGCGCATGGATTCTCACTTTCAGCACACGTTATCAAGAT
GCTTGGAATGATCTTGTGCAAGAGGGCTTAATTTCGAGCGAAAAACGGGATACGTTCAACAT
CCCGATATATACGCCCAGCCTAGAGGAGTTCAAAGAGGTGGTAGAAAGAGATGGTGCATTCA
TAATCAACAAGCTCCAACTTTTCCACGGTGGCAGCGCTCTCATCATCGATGATCCCAACGAT
GCGGTTGAGATTAGCCGTGCCTATGTCAGCCTCTGTCGCAGCCTCACCGGAGGCTTAGTTGA
TGCCCACATAGGCGATCAGCTCGGCCATGAGCTCTTCTCGCGCTTATTAAGCCAAGCCGTGG
ATCAGGCTAAGGAGCTAATGGACCAGTTTCAGCTCGTCCATATAGTTGCATCCCTTACTTTA
GCT

SEQ ID NO:12 – ObCCMT1 amino acid sequence (*Ocimum basilicum* cinnamate/p-coumarate carboxyl methyltransferases; ABV91100)

MARKENYVVSNMNVESVLCMKGGKGEDSYDNNSKMQEQHARSVLHLLMEALDGVGLSSVAAG
AFVVADLGCSSGRNAINTMEFMINHLTEHYTVAAEEPPEFSAFFCDLPSNDFNTLFQLLPPS
DGSSGSYFTAGVAGSFYRRLFPAKSVDFFYSAFSLHWLSQIPKEVMEKGSAAYNEGRVTING
AKESTVNAYKKQFQSDLGVFLRSRSKELKPGGSMFLMLLGRTSPDPADQGAWILTFSTRYQD
AWNDLVQEGLISSEKRDTFNIPIYTPSLEEFKEVVERDGAFIINKLQLFHGGSALIIDDPND
AVEISRAYVSLCRSLTGGLVDAHIGDQLGHELFSRLLSQAVDQAKELMDQFQLVHIVASLTL
A

SEQ ID NO:13 – RgC2'H polynucleotide sequence

ATGGCACCAACCAAAGATTCAGTTATTCACATGGGAGCAGAGTCCTGGGATGAGATTTCCGA
GTTCGTTACTAAAAAGGGACACGGTGTTAAGGGTCTTTCTGAACTTGGTATTAAAACTCTTC
CAAAGCAATTCCATCAGCCTCTTGAAGAGAGGTTCAGTGAGAAAAAGATTTTGGAAAGAGCT
TCAATCCCACTTATCGATATGAGTAAGTGGGACTCCCCTGAGGTTGTGAAGTCTATCTGTGA
TGCTGCAGAACATTGGGGTTTCTTTCAAATAGTTAATCACGGAGTGCCATTGGAGACTTTAC
AGAGAGTTAAAGAAGCTACACATAGGTTTTTCGCTTTGCCTGCAGAAGAGAAAAATAAGTAC
TCTAAGGAAAACTCACCAATTAATAACGTTAGATTCGGTTCTTCATTCGTTCCTCATGTTGA
GAAAGCACTTGAATGGAAGGATTTTCTTAGTATGTTCTATGTTTCCGAAGAGGAAACTAACA
CATACTGGCCACCTATTTGTAGAGACGAGATGTTAGAATACATGAGGAGTTCCGAGGTTCTT
ATCAAAAGATTGATGGAAGTGTTAGTTGTGAAGGGTCTTAAAGTTAAGCAAATCGATGAGAT
AAGAGAACCAATGTTGGTGGGATCAAGAAGAATTAATTTGAACTACTACCCTAAATGCCCAA
ATCCTGAACTTACATTGGGTGTTGGAAGGCATAGTGATATTTCCACCTTTACTATCTTGTTA
CAAGACGAAATCGGTGGACTTCATGTTAGAAAGTTGGATGACACTGGTAACACCTGGGTTCA
TGTTACCCCAATATCTGGTTCACTTATTATCAATATCGGAGATGCTTTGCAGATAATGTCTA
ACGGAAGGTACAAGTCAATAGAACACATGGTTGTGGCAAATGGAACACAAGACAGAATCTCT
GTTCCTTTATTTGTGAACCCAAAGCCTCAGGCTATACTTTGTCCATTCCCTGAGGTTTTGGC
AAATGGAGAAAAACCAGTTTATAAGCCTGTGTTGTGCTCTGATTACTCAAGGCATTTCTACA
CAAAACCTCACGATGGTAAAAAGACAGTGGATTTCGCATTGATGAAC

SEQ ID NO:14 – RgC2'H amino acid sequence (*Ruta graveolens* 2-oxoglutarate-dependent dioxygenase; Vialart *et al.* plant J 2012, 70:460-470)
MAPTKDSVIHMGAESWDEISEFVTKKGHGVKGLSELGIKTLPKQFHQPLEERFSEKKILERA
SIPLIDMSKWDSPEVVKSICDAAEHWGFFQIVNHGVPLETLQRVKEATHRFFALPAEEKNKY
SKENSPINNVRFGSSFVPHVEKALEWKDFLSMFYVSEEETNTYWPPICRDEMLEYMRSSEVL
IKRLMEVLVVKGLKVKQIDEIREPMLVGSRRINLNYYPKCPNPELTLGVGRHSDISTFTILL
QDEIGGLHVRKLDDTGNTWVHVTPISGSLIINIGDALQIMSNGRYKSIEHMVVANGTQDRIS
VPLFVNPKPQAILCPFPEVLANGEKPVYKPVLCSDYSRHFYTKPHDGKKTVDFALMN

SEQ ID NO:15 – Plastid targeting signal polynucleotide sequence
ATGGCTTCGATCTCCTCCTCAGTCGCGACCGTTAGCCGGACCGCCCCTGCTCAGGCCAACAT
GGTGGCTCCGTTCACCGGCCTTAAGTCCAACGCCGCCTTCCCCACCACCAAGAAGGCTAACG
ACTTCTCCACCCTTCCCAGCAACGGTGGAAGAGTTCAATGCATGCAGGTGTGGCCGGCCTAC
GGCAACAAGAAGTTCGAGACGCTGTCGTACCTGCCGCCGCTGTCGACGATGGCGCCCACCGT
GATGATGGCCTCGTCGGCCACCGCCGTCGCTCCGTTCCAGGGGCTCAAGTCCACCGCCAGCC
TCCCCGTCGCCCGCCGCTCCTCCAGAAGCCTCGGCAACGTCAGCAACGGCGGAAGGATCCGG
TGCATGCAG

SEQ ID NO:16 – Plastid targeting signal amino acid sequence
MASISSSVATVSRTAPAQANMVAPFTGLKSNAAFPTTKKANDFSTLPSNGGRVQCMQVWPAY
GNKKFETLSYLPPLSTMAPTVMMASSATAVAPFQGLKSTASLPVARRSSRSLGNVSNGGRIR
CMQ

SEQ ID NO:17 – IRX5 promoter polynucleotide sequence
ATGAAGCCATCCTCTACCTCGGAAAAACTTGTTGCGAGAAGAAGACATGCGATGGCATGGAT
GCTTGGATCTTTGACATTGATGACACTCTTCTCTCAACCATTCCTTACCACAAGAGCAACGG
TTGTTTCGGGTAAATAAACTAAACTTAACCATATACATTAGCCTTGATTCGGTTTTTGGTTT
GATTTATGGATATTAAAGATCCGAATTATATTTGAACAAAAAAAAATGATTATGTCACATAA
AAAAAAATTGGCTTGAATTTTGGTTTAGATGGGTTTAAATGTCTACCTCTAATCATTTCATT
TGTTTTCTGGTTAGCTTTAATTCGGTTTAGAATGAAACCGGGATTGACATGTTACATTGATT
TGAAACAGTGGTGAGCAACTGAACACGACCAAGTTCGAGGAATGGCAAAATTCGGGCAAGGC
ACCAGCGGTTCCACACATGGTGAAGTTGTACCATGAGATCAGAGAGAGAGGTTTCAAGATCT
TTTTGATCTCTTCTCGTAAAGAGTATCTCAGATCTGCCACCGTCGAAAATCTTATTGAAGCC

GGTTACCACAGCTGGTCTAACCTCCTTCTGAGGTTCGAATCATATTTAATAACCGCATTAAA
CCGAAATTTAAATTCTAATTTCACCAAATCAAAAAGTAAAACTAGAACACTTCAGATAAATT
TTGTCGTTCTGTTGACTTCATTTATTCTCTAAACACAAAGAACTATAGACCATAATCGAAAT
AAAAACCCTAAAAACCAAATTTATCTATTTAAAACAAACATTAGCTATTTGAGTTTCTTTTA
GGTAAGTTATTTAAGGTTTTGGAGACTTTAAGATGTTTTCAGCATTTATGGTTGTGTCATTA
ATTTGTTTAGTTTAGTAAAGAAAGAAAAGATAGTAATTAAAGAGTTGGTTGTGAAATCATAT
TTAAAACATTAATAGGTATTTATGTCTAATTTGGGGACAAAATAGTGGAATTCTTTATCATA
TCTAGCTAGTTCTTATCGAGTTTGAACTCGGGTTATGATTATGTTACATGCATTGGTCCATA
TAAATCTATGAGCAATCAATATAATTCGAGCATTTTGGTATAACATAATGAGCCAAGTATAA
CAAAAGTATCAAACCTATGCAGGGGAGAAGATGATGAAAAGAAGAGTGTGAGCCAATACAAA
GCAGATTTGAGGACATGGCTTACAAGTCTTGGGTACAGAGTTTGGGGAGTGATGGGTGCACA
ATGGAACAGCTTCTCTGGTTGTCCAGTTCCCAAGAGAACCTTCAAGCTCCCTAACTCCATCT
ACTATGTCGCCTGATTAAATCTTATTTACTAACAAAACAATAAGATCAGAGTTTCATTCTGA
TTCTTGAGTCTTTTTTTTCTCTCTCCCTCTTTTCATTTCTGGTTTATATAACCAATTCAAAT
GCTTATGATCCATGCATGAACCATGATCATCTTTGTGTTTTTTTTTCCTTCTGTATTACCAT
TTTGGGCCTTTGTGAAATTGATTTTGGGCTTTTGTTATATAATCTCCTCTTTCTCTTTCTCT
ACCTGATTGGATTCAAGAACATAGCCAGATTTGGTAAAGTTTATAAGATACAAAATATTAAG
TAAGACTAAAGTAGAAATACATAATAACTTGAAAGCTACTCTAAGTTATACAAATTCTAAAG
AACTCAAAAGAATAACAAACAGTAGAAGTTGGAAGCTCAAGCAATTAAATTATATAAAAACA
CTAACTACACTGAGCTGTCTCCTTCTTCCACCAAATCTTGTTGCTGTCTCTTGAAGCTTTCT
TATGACACAAACCTTAGACCCAATTTCACTCACAGTTTGGTACAACCTCAGTTTTCTTCACA
ACAAATTCAAACATCTTACCCTTATATTACCTCTTTATCTCTTCAATCATCAAAACACATAG
TCACATACATTTCTCTACCCCACCTTCTGCTCTGCTTCCGAGAGCTCAGTGTACCTCGCC

SEQ ID NO:18 – AtC4H promoter polynucleotide sequence

CGGAATGAGAGACGAGAGCAATGTGCTAAGAGAAGAGATTGGGAAGAGAGAAGAGAAGATAA
AGGAAACGGAAAAGCATATGGAGGAGCTTCACATGGAGCAAGTGAGGCTGAGAAGACGGTCA
AGTGAGCTTACGGAAGAAGTGGAAAGGACGAGAGTGTCTGCATCGGAAATGGCTGAGCAGAA
AAGAGAAGCTATAAGACAGCTTTGTATGTCTCTTGACCATTACAGAGATGGGTACGACAGAC
TTTGGAGAGTTGTTGCAGGACATAAGAGTAAGAGAGTAGTGGTCTTATCAACTTGAAGTGTA
AGAACAATGAGTCAATGACTACGTGCAGGACATTGGACATACCGTGTGTTCTTTTGGATTGA
AATGTTGTTTCGAAGGCTGTTAGTTGATGTTGAAAATAGGTTGAAGTTGAATAATGCATGT
TGATATAGTAAATATCAATGGTAATATTTTCTCATTTCCCAAAACTCAAATGATATCATTTA
ACTATAAACTAACGTAAACTGTTGACAATACACTTATGGTTAAAAATTTGGAGTCTTGTTTT
AGTATACGTATCACCACCGCACGGTTTCAAAACCACATAATTGTAAATGTTATTGGAAAATA
GAACTCGCAATACGTATTGTATTTTGGTAAACATAGCTCTAAGCCTCTAATATATAAGCTCT
CAACAATTCTGGCTAATGGTCCCAAGTAAGAAAAGCCCATGTATTGTAAGGTCATGATCTCA
AAAACGAGGGTGAGGTGGAATACTAACATGAGGAGAAAGTAAGGTGACAAATTTTTGGGGCA
ATAGTGGTGGATATGGTGGGGAGGTAGGTAGCATCATTTCTCCAAGTCGCTGTCTTTCGTGG
TAATGGTAGGTGTGTCTCTCTTTATATTATTTATTACTACTCATTGTAAATTTCTTTTTTCT
ACAATTTGTTTCTGACTCCAAAATACGTCACAAATATAATACTAGGCAAATAATTATTTTAT
TATAAGTCAATAGAGTGGTTGTTGTAAAATTGATTTTTTGATATTGAAAGAGTTCATGGACG
GATGTGTATGCGCCAAATGGTAAGCCCTTGTACTGTGCCGCGCGTATATTTTAACCACCACT
AGTTGTTTCTCTTTTTCAAAAAACACAAAAAAAAAATAATTTGTTTTCTTAACGGCGTCAAA
TCTGACGGCGTCTCAATACGTTCAATTTTTTTCTTTCTTTCACATGGTTTCTCATAGCTTTG
CATTGACCATAGGTAAAGGGATAAGGATAATGGTTTTTTCTCTTGTTTGTTTTATCCTTATT
ATTCAAAAAGGATAAAAAAACAGTGATATTTAGATTTCTTTGATTAAAAAAGTCATTGAAAT
TCATATTTGATTTTTTGCTAAATGTCAACACAGAGACACAAACGTAATGCACTGTCGCCAAT
ATTCATGGATCATGACAATAAATATCACTAGAATAATTAAAAATCAGTAGAATGCAAACAAA
GCATTTTCTAAGTAAAACAGTCTTTTATATTCACGTAATTGGAATTTCCTTTTTTTTTTTTT
GTCGTAATTGGAATTTCCTTTATCAAACCCAAAGTCCAAAACAATCGGCAATGTTTTGCAAA
ATGTTCAAAACTATTGGCGGGTTGGTCTATCCGAATTGAAGATCTTTTCTCCATATGATAGA
CCAACGAAATTCGGCATACGTGTTTTTTTTTTTGTTTTGAAAACCCTTTAAACAACCTTAAT

TCAAAATACTAATGTAACTTTATTGAACGTGCATCTAAAAATTTTGAACTTTGCTTTTGAGA
AATAATCAATGTACCAATAAAGAAGATGTAGTACATACATTATAATTAAATACAAAAAAGGA
ATCACCATATAGTACATGGTAGACAATGAAAAACTTTAAAACATATACAATCAATAATACTC
TTTGTGCATAACTTTTTTTGTCGTCTCGAGTTTATATTTGAGTACTTATACAAACTATTAGA
TTACAAACTGTGCTCAGATACATTAAGTTAATCTTATATACAAGAGCACTCGAGTGTTGTCC
TTAAGTTAATCTTAAGATATCTTGAGGTAAATAGAAATAGTTGACTCGTTTTTATCTTCTTC
TTTTTTTACCATGAGCAAAAAAGATGAAATAAGTTCAAAACGTGACGAATCTATATGTTACT
ACTTAGTATGTGTCAATCATTAAATCGGGAAAACTTCATCATTTCAGGAGTATTACAAAACT
CCTAAGAGTGAGAACGACTACATAGTACATATTTTGATAAAAGACTTGAAAACTTGCTAAAA
CGAATTTGCGAAAATATAATCATACAAGTGCCAGTGATTTTGATCGAATTATTCATAGCTTT
GTAGGATGAACTTAATTAAATAATATCTCACAAAAGTATTGACAGTAACCTAGTACTATACT
ATCTATGTTAGAATATGATTATGATATAATTTATCCCCTCACTTATTCATATGATTTTTGAA
GCAACTACTTTCGTTTTTTTAACATTTTCTTTTGTTGGTTATTGTTAATGAGCATATTTAGT
CGTTTCTTAATTCCACTGAAATAGAAAATACAAAGAGAACTTTAGTTAATAGATATGAACAT
AATCTCACATCCTCCTCCTACCTTCACCAAACACTTTTACATACACTTTGTGGTCTTTCTTT
ACCTACCACCATCAACAACAACACCAAGCCCCACTCACACACACGCAATCACGTTAAATTTA
ACGCCGTTTATTATCTCATCATTCACCAACTCCCACGTACCTAACGCCGTTTACCTTTTGCC
GTTGGTCCTCATTTCTCAAACCAACCAAACCTCTCCCTCTTATAAAATCCTCTCTCCCTTCT
TTATTTCTTCCTCAGCAGCTTCTTCTGCTTTCAATTACTCTCGCC

SEQ ID NO:19 – AtC3H promoter polynucleotide sequence

ATCGTAAGTTTTTTTGTGTGTGTGTTAACAATGTACTCACTACTCACTGTTCCATATTTTTG
ATGTACGTATATCGAAAACATTCTGCCAACAAATGCAAACATAACAAAAGTCAAAAACAATA
ACATAACCGGGAATTAAACCAAAATGTAATTGCTTTTTATTAGTGTCAGGCCTTCTGCTTAA
AAATATTCTCGGCCCAGAGCCCATTAACACCTATCTCAATTCATATTGAAGAAAATGACTAT
ATTACTTGACAAAAACTTTAGTCAGAAAAATATGGAATCTCTTTCGGTACTGCTAAGTGCTA
ACCTTAAATAGTATAGAATTCTTAGTTCATTCTCAAAAACATAGCTATATGTAGATTATAAA
AGTTCGATATTATTTCCTGCAAAAGATGTTATAATGTTACAACTTACAAGAAAATGATGTAT
ATGTAGATTTTATAAACTGGTACCGTAATTCATAAAAGATGGTGGTGGGTATGTATCAGTAA
CGGAACTTACATATGCGTGTGTATTACTATGTCTATATGGTGTATTCCTTTGTGTGGAACAA
TGCACGTCAGAGTTGTTTATTTTCTTATAGAATTTAAGGAATCAATTATTGGATTTCTCAAG
GTGAAAGTGGACTTCTTTGCACGCAAGGTCTAGTTGCCGACTTGCCGTTGCATGTAACATGA
TTGTTGAAATAAAGTGAATTGAGAGAAGTTTGGCCAGACATTTTAAATTTAACCCAAAAAAA
GTAGGGCCTAACACAAAATATAACCTCTCTTTGTTCAAAGGAAATAACACCTACGTCTTATA
ATTGAACCAAACATTGAATCATTGAACTCACCTATAATAATTATAATAACACGAATTCACAA
GACACCTAAAAGAAAAAGTTCACAAAAACAAATAAAAAATTTACCTCTCACCAAACACACTCA
CCTACCCGTCTGGTCCCACTGACCCCAACATACAACACCGACTCTCTCCCACACCAATTTTT
TTTTTTGGCGTTTTAAAACAAATAAACTATCTATTTTTTTTTCTTACCAACTGATTAATTCG
TGAATAATCTATTATCTTCTTCTTTTTTTTGTGACGGATGATTAGTGCGTGGGGAAATCAAA
ATTTACAAAATTTGGGATGATTCCGATTTTTGCCATTCGATTAATTTTGGTTAAAAGATATA
CTATTCATTCACCAAGTTTTCAGATGAGTCTAAAAGATAATATCATTTCACTAGTCACTTAA
AAAAAGGGTTAAAAGAACATCAATAATATCACTGGTTTCCTTAGGTGACCCAAAAAAAGAAG
AAAAAGTCACTAGTTTCTTTTTGGAAATTTTACTGGGCATATAGACGAAGTTGTAATGAGTG
AGTTTAAATTTATCTATGGCACGCAGCTACGTCTGGTCGGACTATACCAAGTTACCAACTCT
CTCTACTTCATGTGATTGCCAATAAAAGGTGACGTCTCTCTCTCTCTCACCAACCCCAAACC
ACTTTCCCCACTCGCTCTCAAAACGCTTGCCACCCAAATCTATGGCTTACGGGGACATGTAT
TAACATATATCACTGAGTGAAAAGAAGGGTTTATTACCGTTGGACCAGTGATCAAACGTGTT
TTATAAAAATTTGGAATTGAAAACATGATTTGACATTTTTAATGATGGCAGCAGACGAAACC
AACAACACTAAGTTTAACGTTCGTGGAGTATACTTTTCTATTTTCGAAGAAGACATATAACT
AAGCTGATTGTTATTCTTCATAGATTTCTTTTCACTGCGAATAAAAGTTTGTGAACATGTCA
CCGTTTGAACACTCAACAATCATAAGCGTTTTACCTTTGTGGGGTGGAGAAGATGACAATGA
GAAAGTCGTCGTACATATAATTTAAGAAAATACTATTCTGACTCTGGAACGTGTAAATAATT
ATCTAAACAGATTGCGAATGTTCTCTACTTTTTTTTTGTTTACATTAAAAATGCAAATTTTA

TAACATTTTACATCGCGTAAATATTCCTGTTTTATCTATAATTAATGAAAGCTACTGAAAAA
AAACATCCAGGTCAGGTACATGTATTTCACCTCAACTTAGTAAATAACCAGTAAAATCCAAA
GTAATTACCTTTTCTCTGGAAATTTTCCTCAGTAGTTTATACCAGTCAAATTAAAACCTCAA
ATCTGAATGTTGAAAATTTGATATCCAAGAAATTTTCTCATTGGAATAAAAGTTCAATCTGA
AAATAGATATTTCTCTACCTCTGTTTTTTTTTTTCTCCACCAACTTTCCCCTACTTATCACT
ATCAATAATCGACATTATCCATCTTTTTTATTGTCTTGAACTTTGCAATTTAATTGCATACT
AGTTTCTTGTTTTACATAAAAGAAGTTTGGTGGTAGCAAATATATATGTCTGAAATTGATTA
TTTAAAAACAAAAAAAGATAAATCGGTTCACCAACCCCCTCCCTAATATAAATCAAAGTCTC
CACCACATATATCTAGAAGAATTCTACAAGTGAATTCGATTTACACTTTTTTTGTCCTTTT
TTATTAATAAATCACTGACCCGAAAATAAAAATAGAAGCAAAACTTC

SEQ ID NO:20 – AtHCT promoter polynucleotide sequence
TTCTCTAGGTTTTGAAGCTTTCCTAGTTCTTTTGGAAGCGTGCCGGACAAGTCATTGTCGTA
TAGAAACAGATTGATAAGTTCAGAGCAGTTTCCAAGCTCTTTAGGGATCTCACCTGAGAGCA
TTGTAGAATAGACAGATAAAGACTGGAGCTTGCTTAGTTGACCCAACGAAACAGGTAAAGAA
CCGGATATTTTCGTTGCTGCTAACCCTAAGACCTTGAGATTCCTACAGTTTCCGATCTCCTC
CGGGATCTTCCCTGAAAGCTCTGAGTTTCCTCCGGCTCTTATGCTCTCAAGAGTCGAGATCT
TTCCGAGCTCCAACGGGAGATTCTCGGATAAGTAGTTATCGAAAATCTCAAGATTCTTGAGG
CTAACGCAGTCGCCGAGTTCCGGTGGGATCTTTCCTGTGAGGCCATTGGAGTTTAAACAAAG
TTCTTGAAGATTCTTGAGCTTCCCTAGACTCGAAGGTATTTCACCAACAAGACTATTTGAGC
TTAAATCGATAACTATAAGCTCCGAACAATCTCCGATCTCAGAAGATATAGCTCCGGTGAGA
TTAGTGTTGGAGATAACGAGTTTCTGAAGTGAAGTAAACGAAGAAATGTTAGGAGGGAAAGG
TAAAGCTAACTGAACAGAGACGACATTGATCTCTGTAACGAGTTTGTTGTCTGAGGAGGAAC
AAGTAATGTAAGGCCATTGACATGGGTCAGAATCAGAAGGATTCCAGCCGGAGAAGACTGAC
GGTGGCGGCGAGTTCGAGCTGTGAAGCCAAGAAATCAAAGCTGAGACTTCATTGGTTGATGC
AGAGGTCGAGGAGATGAAGAAAGCTAAAAACAGAGACAATGTAATGGAAAAATGAGAAACAG
TTAAGGCTTTTTTTCTTGGAATCGGCATTTGCAAAGACATAAGAGTTTTTTTCTTTGCATTT
GGCTCTCAAATCCAAAACAAGCCTTCTTGGTTCTGCATCGATCTGAGTCCTTTGGCTTAGGG
TTTAGGGAAGTTTTTGCTTTAGAGATAAGCAATAAGAAAGAATGATATATTAAATATATAAA
AGTACTAAACTTCATGTGCTCTGTCTTTTTCTTTTACCTCGGGGTTCTGTTTCTAGCTTCAG
ATTAATTAATTACAGTCATTAACTTTTCTTTGAAATATGTTTGCCAAGAGCCCGAGACACTA
TCCATAGATGACAAAAGTCAATAGTTATATATACATAAAATATCACAAAACAAAAGGCATTG
GTTATATATATACAGAATCATTTCACTTAGTAGTGTTTTTTCTTATAAGATTATGATAGAAA
TATGGAAGCATGCATGTGGTTTTGCATTGTTTTCCTCAATTAAGTCAGGATTGTGAGTTGGT
TTGTTTTCGAGACCTGAACCGAGCGTTTAAGATTCTTCCTCGTTTGAAGTAAACTCCATAAT
TGTCCACACCTAAGCTAAAAGAAAGTAATAACAAGTTTAAATATTCATGACAAGGAAAATAT
TGCATTCAGAAAATTGTTAACAACGAAGTAAACATTTTTTTCAATCCGATGCCAATAGTCTC
TAGCGGCATCAAAAGTCCACAAACTCGATACCTCTGGGTAAATGAGCGAATGGGCCGGTCCG
TTGTAGCCCAGAAGAGAAATTGTCCTCTAAATTCCATACTTCCATGAATTTTCTCTGTATAT
CCTCGTTTGATGTATGGTATATTTGTTCCGCTCTAAATCATGACCAACCCAAGGTACTAAAT
TGTCATTTAAGCTTTGATTGGTATTTGGTAGCATGGGTTACCATTGACCAACCCACGGTACT
AGTTGCTTTTCTTTTAGTTTTGCTTTTGCTTTATTTTCTTAGAGAGTGGGAGGACAAAGGT
TTGGATCATTAAGCCAATGAATGCTTCAAAGAAATTGAATTTTTATTAGATCCTCAAACCAA
GTTGGATCATCAAATAATGGCTAAGAAATAATTTTAGAACAGAAAGCAAAGAAAAGCTATCC
GCAACAACAACCATTAGTTAATAAATTAAAATGAAATGTGAAATTTATGACTAATTGAGGTA
TGTTTTCATATAATATAGTATAGTTCGGATATAAATTCAACATAATTTATTTGTGGTGTACT
GAAAAAAAGACTTTCTTGGATTCTGACGTAATTCTCTTAACACGTGAGTTTACGCCGTTAGA
TGTTATTGGTGGTTGTTGTTATGCTCTGCTACGTGGTAATGAGTTAAGTTAAGCCAAACTTT
GGCATTCGATTGACTAACTTGTACGGTAGCTATAACAATCAACTTGTCAATTTTTTTTCCTT
CTTCTTCATTCGAACTTTATACTATTTAAGCCCATTAGTATTATTTGGGCCTTAGGACAGAG
GGAACGGGTTTACCAACCCCGGATAGAAAAGTAGGACCGAGTGATGAGATGGACCAATGATA
AACCTTCTGAGAGAGTTGGTCGACAGATGGAGTAGGCGGGGTCGTGGGGCGGTAGGTGAAGG

ATTACGACCTTTCCTTTTTTGTTCACACCCACTTATATCTACCCCTCCTCGCTTCTCACACA
ATTTCTCAGATCAAACTCAAAACAAAATTTGTTTGTTCGTTTGATCTTTCTTAAAAAT

SEQ ID NO:21 – AtCCR1 promoter polynucleotide sequence
TTGCTTTCTCTGTCCATGATATGAGGCATTGACTTCTCACCTGTATTCATATGGTATAGATTCCTCTT
TTCAGGAGTCCAATACAAACGAGCTTGGTGAAGAACTCGTTGGTAAGAGAGTTAATGTCTGGTGGCCA
CTCGACAAGAAGTAAGTTTATTGTTAAACTTACTAACTTCATTTTTGATACTATATGATGAATGATAG
CAATCTTACGATTTGTATTTGCACAGGTTTTATGAAGGTGTCATAAAATCTTATTGTAGAGTTAAGAA
GATGCATCAAGTGAGTTAACTTCTCTATTTGGTATTTTAAAATTCTCTATTTATTGCATAACTGGTTT
ATATAGAATTTTCCCACTGATGGTCTCGCAGGTAACATATTCTGATGGCGATGTTGAAGAGCTTAATC
TGAAAAAAGAACGTTTTAAGATAATCGAGGATAAATCTTCAGCCAGTGAGGTGAAAATTTCTTACATT
CTATCATTCACCATTCTTTATATTTACCAAAATTTCAATGTATCTGGTTTCCCTAATAAAATCTAAGC
AGGATAAGGAAGATGATCTGCTTGAGTCTACTCCTTTATCTGCCTTGTAAGTGAAACTTCCATAGTTC
TATGATAACCCACAATTTATAATTTTAATTTAGCTTTAGTCTTGAGTTTTTTGCTGTTATGTGCAGTA
TACAAAGGGAGAAATCCAAGAAGAGGAAAATTGTGTCTAAGAATGTGGAACCGAGTAGTTCTCCAGAA
GTCAGGTATGAAAGTATATAAGAATTCTAGTTTTAGTTGTTTGAAAGTTTGATCCGTGAGTGAATTAG
TTCACAATTATGGATGTAGATCCTCTATGCAAACAATGAAGAAGAAAGACTCTGTAACAGACTCCATT
AAGCAAACAAAAAGAACCAAAGGTGCACTGAAGGCTGTAAGCAATGAACCAGAAAGCACTACAGGGAA
AAATCTTAAATCCTTGAAAAAGCTGAATGGTGAACCTGATAAAACAAGAGGCAGAACTGGCAAAAAGC
AGAAGGTGACTCAAGCTATGCACCGGAAAATCGAAAAAGATTGTGATGAGCAGGAAGACCTCGAAACC
AAAGATGAAGAAGACAGTCTGAAATTGGGGAAAGAATCAGATGCAGAGCCTGATCGTATGGAAGATCA
CCAAGAATTGCCTGAAAATCACAATGTAGAAACCAAAACTGATGGAGAAGAGCAGGAGGCAGCGAAAG
AGCCAACGGCAGAGTCTAAAACTAATGGAGAGGAGCCAAATGCAGAACCCGAAACTGATGGAAAAGAG
CATAAATCATTGAAGGAGCCAAATGCAGAGCCCAAATCTGATGGAGAAGAGCAGGAGGCAGCAAAAGA
GCCAAATGCTGAGCTCAAAACTGATGGAGAAAATCAGGAGGCAGCAAAAGAGCTAACTGCAGAACGCA
AAACTGATGAGGAAGAGCACAAGGTAGCTGATGAGGTAGAGCAAAAGTCACACAAACACACAAATCTA
GAACCGGAAGCTGAGGGAGAAGAGCAAAAGTCAGTGGAAGAGCCAAATGCAGAACCCAAGACCAAGGT
AGAAGAGAAAGAGTCAGCAAAAGAGCAAACTGCAGACACAAAATTGATTGAGAAGGAGGATATGTCTA
AGACAAAGGGAGAAGAGATTGATAAAGAAACATATTCAAGCATCCCTGAGACTGGTAAAGTAGGAAAC
GAAGCTGAAGAAGATGATCAGAGAGTGATTAAGGAACTGGAAGAAGAGTCTGACAAGGCAGAAGTCAG
TACTACGGTGCTTGAGGTTGATCCATGAATGAAGGATTGTTAGGTAAATGTTAATCCAGGAAAAAAAG
ATTGGTTCTTGTGGTTTAGGTAACTTATGTATTAAGTGAAGCTGCTTGTTTAGAGACTAATGGTGTGT
TTTATGAGTAGATTCTTCTGACCTATGTCTCGTTATGGAACTAGTTTGATCTTATGTCACCTTGCTAG
CAGCAGATATTGATATTTATATATTTAAGAGACATGCGCATGAGAATGAGGGTATGGAAAAGTCCATA
TCAGATGACACAAACAATGATCGTATGTGTAGTCACTTGTGCATTTCCAGTTTTGGACATAAAATTCT
GATATTGCATAGAAATGTTTTTAAATAACACTAATCCAAACCTAAATAAAATATCTCTATACATCATC
TAGAAATGTATGGCTTGATCAAGAATTGTAGATAATAATACCCTGAGTTAAATGATTGTAGGTATTAT
TTCAGTTTTCAAAATTGTCCAAATTTATGAGCTATATTAAAGATAATATTTTCAATAAGGTGTGTAGT
TCTAAATGTTTCTTCTTCTTCCACCAACCCCTCTTTCTATATGTATGTTCTTTTTTCTAAAATAATTG
TTTGTTCTTTTTTAGATATATCAAATTAAATATAAAAAATATTGACAAAACTTATTTACCATTGTTAG
GTGAACTTGGCAAGTGTGTAAATATAAAGATAACATTCCTTTTCGTTCTTTATATATACGAAACGTAC
CACAAATTTCTAACTAAAGCATTCATAGTCTCTCGAAAGCCTCTTTTCAGAACCGAAGCTCTTTACTT
TCGTCCACCGGGAAAT

SEQ ID NO:22 – AtCAD4 promoter polynucleotide sequence
CAGAAAGGTCTTCACACTCTGTTTTAGCTAGAGAGTTTTATCCATCTGAGTTTTTAGTCTATTTTGTT
TTATCTAGGAGTTGCTTTGTTTGTTCGAATTCGGTCATTGCTTTTGCTGCTTTACTGGAGTCAAATTT
GAAGGTAAAATATATGTTAAATATCTGGGTAGGTGGTTGTGGATGATGGAAAATCTGAACGTATCACT
GTTAATGACAATGGAGAACTCGTTTCTACTCAGCATGCTATCACCGAATACCGAGTGATTGAATCTTC
ACCACATGGTTAGTGAGACTGACTTCCATTTCTATTCAGTTAAACTTAAAGCAAATGATTTTGCCTTG
AGTTTTTAGCACATTGTTGAATTGCAGGATACACATGGCTTGAGCTTCGCCCTTTAACCGGGAGAAAA
CATCAGGTCTCTATAGATATTCAGTTTTTGTTTCAACTTTCTCTCTTTTTTATGTTCTCTTAATACTA
ATCTGTTTTCAACTGTTCTTCGATTGCCACAGCTTCGTGTACACTGCGCTGAAGTGCTAGGAACACCG
ATAGTCGGGGACTACAAATACGGTTGGCAAGCTCATAAAGCCCGGGAACCTTTTGTCTCTTCTGAAAA
CAACCCAACCAAGCAATCATCATCTCCTTTTGGATTGGATCTGGATGGTGGAGATGTCTCTTCGAAAC
AGCCACACCTTCATCTCCATTCAAAGCAAATCGATCTGCCAAACATATCACAGCTCTTGGAGAAAATG
CAGGTCTCTTCAGACTCTGATATTTCGGATCTCGATAGCCTTAAATTCGATGCTCCATTGCCTAGTCA

TATGCAACTAAGCTTTAATTTGTTGAAATCTAGAGTCGAAACTTGTGACAAAAATTAGATTTTTTTTC
TTACCGAGCTTTCTTCTTTGTGTTCATTGAGGCCCAAGTATTTGTGTATTTGGACCTGAATATTCTCA
TACAAAGATAAATAATTATAATTAAATGATTTTTCGCATATAATCATTATTGTGGTATGATTAACACA
GTTGGTGTGATGACTGATTGACACAATAATCACCGTTTGGATTCGATTCCTTTAATACTTGTCACTAG
AGTTGTTTGACTAAACAGCTAACTTGTCACTAGAGTTATTGTGTTTGTATTTTGATCTGTTATTAATC
TGATTGGGTATAATTACAGATAGAGAGACATCTATATTGTAATTAAGACAATCTTAAAGTGTAAACTA
AAAAGATCTCTCTGACCTCTGGAAAACGAAAGGTGGGTGACACATCACTCTAGCTATGAATATGATGA
ATATTCAGTACCTAACCGAACAAAGACTGGTTTGGTATTTTTATTGGAAAAAAGAGATAAATAATTGT
GAATGTGAATTATCCTGTCTGAAAGGTAAGCTGATGACATGGCGTTATATGATTGGACGAGCTTCAGA
ACAAAAGAGTAGCGTCGAATCGAATCTTTACCTACTACACTTTGAACTTTGAAGTACATTACCTACTT
CCTCCTTGATCGAACGTCTTTTCTCAAAACTATTTTATTTCCCCAATTAAAGTAGTGGTGATAAATTC
ACAAAAATACAAACACTTTTATTTTTGACGTCAAAAACAAATACTTCTTTGAACAGGCTATTACAATA
TTTTTAAGAAAAAAGTAAGCAAAATAGTCCACAAACCAAAATCTGTAACATATTAAACGATTTATGTT
TTTTTTTTTTTTTCTTAACTAGAGAACAATTCGGGCTTTTACTAAGGATGATGAGTGTAGTTACCGAA
TAGTGTATTCATATAATCTTTTAATGAGCTTAAGATATGATATTATTTCGACTAATCAGATAAGAGTA
GTTAGATAATTTCGTAATAGAGCAACTCTTTCGCAAATAAAACCATTGTAAACATTACCAATTAGTTT
TTCTTTTTTTTTGGTCACAACCAATTAGTTTGTTTGTTCTATTTTATGAAGTGCGTATTAAAGCTAAC
GTGTTTACAGTAACGCCACACAAATAAAAATAAAAATAATTATGTACTTTATGGATTTATAGAAAAAA
CAAGAATAGTCACCAAAAATTGATTGTGTCATATATCTTTTGTCAACTATTTTATCTTATTTTTCTAT
GGATATGTATGTCCAAAATGTTAGACAAAAAACCAAAAAATCATGTCCAAAATTTCGTTAGGCTGCCG
ATATCTCTGTTTCCCTTTCAACGACTATCTATTTAATTACCGTCGTCCACATTGTTTTTAATATCTTT
ATTCGAGGTTGGTTTAGTTTTTTTTACCAAACTCACTTTGCTACGTTTTTGCCTTTTTGGTATGGTTG
TATTTGTACCACCGGGAAAAAAAAGATAAGAGGTTTGGTTGGTCGAGCTTACTGATTAAAAAATATAC
ACGTCCACCAAATATTAAAACAATATATCCCATTTTTCCTCCTCTCTTTTGGTATTACATTAATATTT
TATTATTTCCCCATTTGCTCTGTATATATAAACATATGTCAATAGAGTGCCTCTACAGTCATGTTTCC
ATAGACATAATCTCTCACCATTGTTTTTCTCTGCAAAACTAAAGAAACAAAAAAAGAAAAATCGGAGA
AACCAAGAAAAAAGAA

SEQ ID NO:23 – AtCAD5 promoter polynucleotide sequence

CCTCGATAACTCTGATTGTTGTATTGTCCAAGTATTCACTAAACAACTTTGCTAAAAGAGAAGATGCT
GCTGGAGCAATTTCAGAAGGTTTTAGCACAACCGCATTACCAGCTGCAATAGCTCCAATGACTGGCTC
GACAGACAATACTAAGGAAAAAAACAAAGCACCATGAAGACATATAAACTTTAATAGTTTAGAAATTG
AGACAAAATTGTCAATAAATAAAATTGAGCTTACAGAAAGGGAAATTCCAGGCTGAAATAACCAAAAC
AACTCCAAGCGGTTCTGAGACTATTTGTGCAGACGAGGGAAATGTTGTCACAGAAGTTTTGACCTGAA
AGGTCCAAGCATAGAAAAAGCAAGTGGTTTTAGAAAGGACACATATCAATGAAGCAGCAAAGCTTGAA
CGGTCTAGTTACCGTTTCTGGAGCCATCCAGTTCTTTAACTCTTTGATTGCAAGCATACAGGATGATT
TTGTATTCGAAATCTAAAAAACGAGAAAAATACCAAAGAGATTCAACAGTGGATAAGTGGAATGCAGT
GAAGAAACGGGACATTGAAATTATATAAAAAACCTCAGCTAGAAAAGCTTCAAGCTCAGGCTTAGAAA
GATCTTGATACAAAGCTTCGGTGATGCATTTCTCCTTCTCATCAATCATCCTAGCAATGTTTTGAAGC
TGAGAAATTCTCCACTCGTAGCTCTTCGTTCTGCCAGAGTTGAAGTTGCTTCTGAGCTCATCTACAAG
CAAAGCTGCTTCTTTTCCACTAAAGTCTGATGCTTGCTCCTTTACCACAGCAGATAGTGTTGCATAAC
AAGTACTGATTCAAGACACCAAAACCGCAATGTGAGAGACTTTAAGACTAAAAATCATGGATAAGACT
AAAAAAACATGGATAAGTATCAACTGTTCTCACGATTATTTATTCATACCACTGTACTTAAACTTAAA
ACCCACTATACTAAATAGAAAGGTAATCATCAAAAAATCAGTATGTAAAAACCACTTTTGTGAATAAA
ATATGTAAAATGGGTGAATAAAGAAATGTGCTTACAATTTCAACCGATAAGGGATACAAGCATTGCTG
CAATATCCACCACCACCACGACGAGATATCCGAAAAGGTGAAGTTGCAACATTTAATCTGCAACAAAA
GAGGCCATTCATTAAAATGGTACTAATTAGATCTAATCATATCATATTGAATGACCAAATCATTCACA
GAAGCATCCATTGCTCCAATTAACATTCTAGACCAAATTCAACTTAAAGGTAACTCTTTTATACAGGA
AACCGAGAAACCGAAAACGCAATTCACATAAAAAGGAAGGCTTGTTTGGAGAAGCAGAATCGAACAAG
TCAATCTCAAACCCTGATGAGCAGGTTTTTCAAGTTACCTGGCAGGAGAAAAACCCTTGGCAAAACAA
AGGGTTTGAATATGATTAATCTCTAGAAGCTTCGTCATGACTTGGGTTCAGTTAAAAATCTCAAATTG
GAGACATTATTGGTGTTTATATATTTGAGAGAGAGAGCCAGAGAGGAGACGTTGAATTGAATGAAGGG
TGTGGTCGGAAGAGAAGACGTGTAGAAGAGACGAGACAAGTAAATTTAAGCATTGGCCCCATTTACAG
CCACAAGTCCGCTACAACAAATTATTTCCAAGAAACTCTGAGATAACGTCGTGATGAAACGGCTCATG
CTGCTGTTGTGATTCGTGAATTAGAGGTTTATCTTTTGGGTTTTTGAATGTTACTTAATTGGACGGTC
GATTTTTCAAACTGGGTGTGAAATGTGAATGGGTCATTCATAATGGCTTTTGTTTTAATGTGAAGCC
ATTCACACACTCTTTGTCCTTCTTTTCTATTATTCATAACTGTCACTCTTTGTTCTTCGAAATAGTAA

AGAGCAAATCGATTCTTTGTTGATCTGGGCCGTAAAATTTCCATGGTTGTGGGAAGTATTCTCGCAGC
TGATCTGGGCCGTCAATGCTACAGTTTCATGTCAGAGAGAGGTCAAGAATCAACACGTGGCCAACCAT
GATTTTAAACCAAAGCAAACACACGATTAGACCCCACATTGTTTGTTCACCAACCCCCGTGGACCCTC
CTTTAGCCGACGTGTCCACGTCAATAGTGGTTTTTCTTCCTTTCAAAGTACACAAATTCCATTCTTTC
TCATTTTACTTTTTGGATTACGTTGTTGTTATAAACTGGTAAAATGAATTATGAATGCAAATAAATTT
CATTTAAGTTTTGTTGGCTTCTAATATTTTTTTCACCTAAAATTCTAATAAACTACACAGCCATGAGC
CATCGTATGAAAAGAAGAAGAAAAAAAATGTCTTTTTCTAGAAGGATCTTTCAACGACTAAAAAAGAT
TTTAAGCTTTTGACTAATTTTGTCAATAATATACACAAATTTACACTCAATTATAGCCATCAAATGTG
TGCTATGCAGAAACACCAATTATTTCATCACACATACGCATACGTTACGTTTCCAACTTTCTCTATAT
ATATATATAGTAATACACACACATAAACAGCAAAAGCGTGAAAGCAGCAGATCAAGATAAGAAAGAAG
AAAGAATCATCAAAAA

SEQ ID NO:24 – AtF5H promoter polynucleotide sequence
TGTGTGTCTTTTTGCGAGTAGTTGTTGGCTTCAGACAGTTCATAGCGGAGTTACTCTATACGCGAAGT
ACTTGTCTCATACTGATAATTTTGATGGCAATTAAGGCTTTAAAAGCTTATGTATTTTCTTATAACCA
TTTTATTCTGTATATAGGGGGACAGAAACATAATAAGTAACAAATAGTGGTTTTATTTTTTTAAATAT
ACAAAAACTGTTTAACCATTTTATTTCTTGGTTAGCAAAATTTTGATATATTCTTAAGAAACTAATAT
TTTAGGTTGATATATTGCAGTCACTAAATAGTTTTAAAAGACACGAAGTTGGTAAGAACAGGCATATA
TTATTCGATTTAATTAGGAATGCTTATGTTAATCTGATTCGACTAATTAGAAACGACGATACTATGAG
CTCATAGATGGTCCCACGACCCACTCTCCCATTTGATCAATATTCAACTGAGCAATGAAACTAATTAA
AAACGTGGTTAGATTAAAAAAATAAATTGTGCAGGTAGCGGATATATAATACTAGTAGGGGTTAAAAA
TAAAATAAAACACCACAGTATTAAATTTTTGTTTCAAAAGTATTATCAATAGTTTTTTTGCTTCAAAA
ATATCACAAATTTTTGTATGAAATATTTCTTTAACGAAAATAAATTAAATAAAATTTAAAATTTATAT
TTGGAGTTCTATTTTTAATTTAGAGTTTTTATTGTTACCACATTTTTTGAATTATTCTAATATTAATT
TGTGATATTATTACAAAAAGTAAAAATATGATATTTTAGAATACTATTATCGATATTTGATATTATTG
ACCTTAGCTTTGTTTGGGTGGAGACATGTGATTATCTTATTACCTTTTTATTCCATGAAACTACAGAG
TTCGCCAGGTACCATACATGCACACACCCTCGTGAAACGAGCGTGACTTAATATGATCTAGAACTTAA
ATAGTACTACTAATTGTGTCATTTGAACTTTCTCCTATGTCGGTTTCACTTCATGTATCGCAGAACAG
GTGGAATACAGTGTCCTTGAGTTTCACCCAAATCGGTCCAATTTTGTGATATATATTGCGATACAGAC
ATACAGCCTACAGAGTTTTGTCTTAGCCCACTGGTTGGCAAACGAAATTGTCTTTATTTTTTTATGTT
TTGTTGTCAATGTGTCTTTGTTTTTAACTAGATTGAGGTTTAATTTTAATACATTTGTTAGTTTACAG
ATTATGCAGTGTAATCTGATAATGTAAGTTGAACTGCGTTGGTCAAAGTCTTGTGTAACGCACTGTAT
CTAAATTGTGAGTAACGACAAAATAATTAAAATTAAAGGACCTTCAAGTATTATTAGTATCTCTGTC
TAAGATGCACAGGTATTCAGTAATAGTAATAAATAATTACTTGTATAATTAATATCTAATTAGTAAAC
CTTGTGTCTAAACCTAAATGAGCATAAATCCAAAAGCAAAAATCTAAACCTAACTGAAAAAGTCATTA
CGAAAAAAAGAAAAAAAAAAGAGAAAAAACTACCTGAAAAGTCATGCACAACGTTCATCTTGGCTAAA
TTTATTTAGTTTATTAAATACAAAAATGGCGAGTTTCTGGAGTTTGTTGAAAATATATTTGTTTAGCC
ACTTTAGAATTTCTTGTTTTAATTTGTTATTAAGATATATCGAGATAATGCGTTTATATCACCAATAT
TTTTGCCAAACTAGTCCTATACAGTCATTTTTCAACAGCTATGTTCACTAATTTAAAACCCACTGAAA
GTCAATCATGATTCGTCATATTTATATGCTCGAATTCAGTAAAATCCGTTTGGTATACTATTTATTTC
GTATAAGTATGTAATTCCACTAGATTTCCTTAAACTAAATTATATATTTACATAATTGTTTTCTTTAA
AAGTCTACAACAGTTATTAAGTTATAGGAAATTATTTCTTTTATTTTTTTTTTTTTTTAGGAAATTAT
TTCTTTTGCAACACATTTGTCGTTTGCAAACTTTTAAAAGAAAATAAATGATTGTTATAATTGATTAC
ATTTCAGTTTATGACAGATTTTTTTTATCTAACCTTTAATGTTTGTTTCCTGTTTTTAGGAAAATCAT
ACCAAAATATATTTGTGATCACAGTAAATCACGGAATAGTTATGACCAAGATTTTCAAAGTAATACTT
AGAATCCTATTAAATAAACGAAATTTTAGGAAGAAATAATCAAGATTTTAGGAAACGATTTGAGCAAG
GATTTAGAAGATTTGAATCTTTAATTAAATATTTTCATTCCTAAATAATTAATGCTAGTGGCATAATA
TTGTAAATAAGTTCAAGTACATGATTAATTTGTTAAAATGGTTGAAAAATATATATATGTAGATTTTT
TCAAAAGGTATACTAATTATTTTCATATTTTCAAGAAAATATAAGAAATGGTGTGTACATATATGCAT
GAAGAAATTTAAGTAGATAATACAAAAATGTCAAAAAAAGGGACCACACAATTTGATTATAAAACCTA
CCTCTCTAATCACATCCCAAAATGGAGAACTTTGCCTCCTGACAACATTTCAGAAAATAATCGAATCC
AAAAAAAACACTCAAT

SEQ ID NO:25 – AtPAL1 promoter polynucleotide sequence
CAAATAGTACGATGTATTTAGTGATTTTATTTATGTACTTTGTTCATTAAATTAGTCATAATTGTTCT
GATTTTTAGGGGTTTTGATCGAACCCTTAGATCAAAAGTTACCTTAATTGTTTTTTTAGCTAAGTACT
TTATTAAAAATTTAATGTTTAGTTCTGATTGAGTAGTACTATAAAGGAGACATGTGTCAATCTTGTCA

ATTGGTTTTGAGTTCAACAATATGCAATATTGCACATGCATTAACGACCAAAAGAAGATGCAATGCAC
TTAAATCATTGAAACTGATTTTGTTTTTGTAGTGTATAAAATATCTATTTAATTACCAACGAAAGAAG
TGAGCTTTTAAAAACAAAGAGTCAGAAGATATATATAACTACAAAACCTACAGAAGATAAGCTGGATT
TCAAAAGAAGAGAAAGAGTAAACCAATAAATTGACCAAAGCAAAATCGGATATTTGACATAAGTTTCC
ATTCACATTGACCCAAATCCACCAGCATTTCAAATAAAGTTACTTAATATAATTTTTGTGTTTATAAT
ATATTCCGCCCACTCTTGCCTTCATTTGGACCTTATCCTAAAAGTCAAAACAGGTGAAAAAAATGAGA
ATACAATTAACACGAAAAATGCAAAAGACTGTTAAACCGAAATCGAATTCTAGTGTAATCAATCCTTT
TCCCAATGATACAACTATAAATCAAAAAGAAAAAATGTACTGATAAACGAAACTAAACGTATAAATTA
ATATATTTCTTGACATAAATAGGAGGCTTTTGCCTGCTAGTCTGCTACGATGGAAGGAAAAATGCATG
CACACATGACACATGCAAAATGTTTCAATGAAGACGCATTGCCCAATTAACCAACACACCACTTCTTC
CATTCCACCCATATTATTTATTTCTACCATTTTCTTTAATTTATTGTTTTTTCTTTGATTCATACACT
GTTTATGACTATTACATTTTCCCTTTCGACTAATATTAACGCGTTTAAACCAAAGAATGGATTTGATA
ATGAAATTTTATTTTATTAGCATATAGATAATGGATGGCTTCATGCTTGGTTTCCATGACAAGGAATG
ACACAAGATAATTATTTTGAATAAAATCATAAATATGATAATACTAGTTGTAAAAAAACTTGAGTGTT
TCGTGTGTTATTTTTCGGTTTCTTGACTTTTTATATTTCTCGTTTTTGTAATTTTAGGATGGATTATT
TAGCTTGCTTTTCTCTTTTATTACTTTCTAAAATTTTATTTATAAACTCATTTTTAATATATTGACAA
TCAATAAATGAGTTATCTTTTAATTAATAAAAAATTTGTAAACTCTTGTAAACAGATCATAGTCACTA
AAAGCTATTATAAGTTATTTGTAGCTATATTTTTTTATTTCATGAACTTAGGATAAGATACGAAAATG
GAGGTTATATTTACATAAATGTCACCACATTGCCTTTGTCATGCAAACGGCGTGTTGCGTCACTCGCC
TCCTATTGGGAATCTTATAATCGCGTGAATATTATTAGAGTTTGCGATATTTCCACGTAATAGTTATC
TTTCACAAATTTTATACTCAATTACAAAATCAACGAAAATGTACATTTGTATCTTTAACTATTTACGT
TTTTTTTACGTATCAACTTTCAGTTATATGTTTTGGATAATATATTTTTTTACTTTTGACTTTTCAGT
TTTCACCTAATGATTGGGATATACATATGCATGCATAGTTCCCATTATTTAAATGTAAGCTAAGTGCA
TATGAACTGTTAGTCAAAATTACGAAGTTTATTTGTACATATATATAGTTATAACAAAATGGTACAGT
AAATTAAACAGAACATCAAGAAAGTACAAAAGACTGAACACAATAATTTACATGAAAACAAAACACTT
AAAAAATCATCCGATAAAATCGAAATGATATCCCAAATGACAAAAATAACAATATAGAAAATACAAAA
ACAAAAACAAAATATGAAAGAGTGTTATGGTGGGACGTTAATTGACTCAATTACGTTCATACATTAT
ACACACCTACTCCCATCACAATGAAACGCTTTACTCCAAAAAAAAAAAAAAAACCACTCTTCAAAAAA
TCTCGTAGTCTCACCAACCGCGAAATGCAACTATCGTCAGCCACCAGCCACGACCACTTTTACCACCG
TGACGTTGACGAAAACCAAAGAAATTCACCACCGTGTTAAAATCAAATTAAAAATAACTCTCTTTTTG
CGACTTAAACCAAATCCACGAATTATAATCTCCACCACTAAAATCCATCACTCACTCTCCATCTAACG
GTCATCATTAATTCTCAACCAACTCCTTCTTTCTCACTAATTTTCATTTTTTCTATAATCTTTATATG
GAAGAAAAAAAGAAACTAGCTATCTCTATACGCTTACCTACCAACAAACACTACCACCTTATTTAAAC
CACCCTTCATTCATCTAATTTTCCTCAGGAACAAATACAATTCCTTAACCAACAATATTACAAATAAG
CTCCTATCTTCTTTCTTTCTTTTAGAGATCTTGTAATCTCCTCTTAGTTAATCTTCTATTGTAAAACT
AAGATCAAAAGTCTAA

SEQ ID NO:26 – AtPAL2 promoter polynucleotide sequence

GATTGATGGTTTAATAATCTGCCTCGTGATACATGGTGTTATCTTAAAATGGTCTCTCAATTAGTCTT
TGTATTTGTATAAAATAAGGCCTAAAAATATCATCAATGGGGTCCTGTTAAAAACAAAAACAGATACA
CCTTTCACTAATAAAAAAAAACTGTTACCGACAAGTCAAACAATATCTGCGGACAAAAAAATGAAGAA
TGTTTAGTAAGAAATAGAAGATGTGGTAAAGAGCCATACACACATGCAAGTGTTTTTCAATGAACCCA
TCTTACCAACCCACTACTTCTTTGAGCCATAATTGTTTGGTTCGGAGACCCTTTACATTTCCGTCTCA
GCTTTATTTGTTTACGCATTGATTTGTCTTAAATTATGTTAGATATTGTTTTTTGGCTATTTATTAGC
AGCAATCAAGTTAAAAGAGTGGTTCGATATCACCATCGAACTCTCCTTTAGATATTTTCTATATAAAA
CCAAACAAAAACAAAAAAATTGGTCCGATCATCTAATATACAAGTTAGACGATTTCACGTTATGTTAT
TACAACCTACAACAAAATAGACTATGATCGAAATCATATTGAATCTTTTACCTTTCAACGTAATACAA
ATCTGGCTTTACAAAGCAATAATTCATGTTTGTTTGTCTAATTTAAATTTCCCTGTTTTTTTTCCCCT
CTTTCTGTTTCCCATTTGAAAGTAAAAGATCATTTAAGCACCTAACTCAATTTTATTTTATTTTAAAC
ACCTAATGTCATGCTCCTTGGCTCCTTGTAATTAGTTGATCGTTTCAATTTAGACCAGCAAAACATTT
TAGTATGTTCGTAAATATTGCGTACATGCCATTTCGTTTGTCATGCAAACGGTGTGTGTTTCTTTACT
TAGCTTCTAGTTGGTGTATATTGCGTCGCATTAATATCGGTTTACCTTCCTCCTGTCTACGTAATGAT
ATATTCTCCACCACAAATTTAAATTCTTATTGAAATTTCCTAATTTTTTAGGTAGCTCAAGGTCTCAA
GTATACTACGTACCCTATTTTTTTGAATATCTATCTATATTATAACAAGAGTTTTTCTGAGCTAGTTA
ATGAGATGACAATATTCTACATAAATAAATGACCCTCGAAAGTTTCAAGTACTTTAGGATCTGACCAA
ATCGGGGTAAAACATTTTGAAACTAATTACGTTCACATCTACCATCGATGATTGACAAGCTTATTGTC
ACCTTTTATGTTAAAGTGACATGGTCTTGACGTTAATTTGCATGTTATTCTACATCTATAGTCCAAAG

ATAGCAAACCAAAGAAAAAAATTGTCACAGAGGGTTCAATGTTACTTAGATAGAAATGGTTCTTTACA
ATAATAAATTTATGTTCCATTCTTCATGGACCGATGGTATATATATGACTATATATATGTTACAAGAA
AAACAAAAACTTATATTTTCTAAATATGTCTTCATCCATGTCACTAGCTCATTGTGTATACATTTACT
TGCTTCTTTTTGTTCTATTTCATTTCCTCTAACAAATTATTCCTTATATTTTGTGATGTACTGAATTA
TTATGAAAAAAAACCTTTACACTTGATAGAGAAGCATATTTGGAAACGTATATAATTTGTTTAATTGG
AGTCACCAAAATTATACAAATCTTGTAATATCATTAACATAATAGCAAACTAATTAAATATATGTTTT
GAGGTCAAATGTTCGGTTTAGTGTTGAAACTGAAAAAAATTATTGGTTAATAAAATTTCAAATAAAAG
GACAGGTCTTTCTCACCAAAACAAATTTCAAGTATAGATAAGAAAAATATAATAAGATAAACAATTCA
TGCTGGTTTGGTTCGACTTCAACTAGTTAGTTGTATAAGAATATATTTTTTTAATACATTTTTTTAGC
AACTTTTGTTTTTGATACATATAAACAAATATTCACAATAAAACCAAACTACAAATAGCAACTAAAAT
AATTTTTTGAAAACGAAATTAGTGGGACGACCTTGAATTGACTGAACTACATTCCTACGTTCCACAA
CTACTCCCATTTCATTCCCAAACCATAATCAATCACTCGTATAAACATTTTTGTCTCCAAAAAGTCTC
ACCAACCGCAAAACGCTTATTAGTTATTACCTTCTCAATTCCTCAGCCACCAGCCACGACTACCTTTT
CGATGCTTGAGGTTGATATTTGACGGAACACACAAATTTAACCAAACCAAACCAAAACCAAACGCGTT
TTAAATCTAAAAACTAATTGACAAACTCTTTTTGCGACTCAAACCAAATTCACGTTTTCCATTATCCA
CCATTAGATCACCAATCTTCATCCAACGGTCATCATTAAACTCTCACCCACCCCTCATACTTCACTTT
TTTTCTCCAAAAAATCAAAACTTGTGTTCTCTCTTCTCTCTTCTCTTGTCCTTACCTAACAACAACAC
TAACATTGTCCTTCTTATTTAAACGTCTCTTCTCTCTTCTTCCTCCTCAGAAAACCAAAAACCACCAA
CAATTCAAACTCTCTCTTTCTCCTTTCACCAAACAATACAAGAGATCTGATCTCATTCACCTAAACAC
AACTTCTTGAAAACCA

SEQ ID NO:27 – At4CI1 promoter polynucleotide sequence

ACATAAGATTTGGATTATGAGAGGAGTTGAGAAGTTATATGATGGAAACTGAAAAGTAAATCTTTTTG
CAGAGCTGTAGAATCAATCAACATTTGATGACTTGGACTTCTTCACCATGTGTGTTGGTGTGGACCAT
TGAATTGACGGTTTTGCCATTCACCAACAACAGCATGAGTTTTTGAGTCTTCATGTTTGGTAAAGGTT
AGGCTTATTAGGAGACACGGGTAAGAGACTAGAGAGAGACATTCTCCAAACCTTTCTTTTGCATGTTT
TGTAAGAAACATTTCCGAAAATGAAAGAAATCTTACACAACATTCATATAATTTGTTTGAAATATAAC
AAAATGATAATTTATACTCTCAAGTAAAATGCCTAAACTTTTATCAATTGGAAAAGACATCACACACA
AGCGTGAAGCGTATCTTATTACCAAACCCAACTAAGCATGGGTCTCGATACTTGCCATAATTACTTTA
ATCCATTCTCTTTTTGAGAAATGTATAAAACATGACTTTGCATAAATAGTCTTTTACTAATTACTATG
TAAATAATTCCTAAGACTGGTTTCATGGTACATATTATCGTTTTATCCTTGTTTTAAGAATATTCAGA
TGTTTGGTCTATGGAATATAGTCTATTCTTCATGTTTAAAACTATTATTTGATAAGAAAATATGTACT
AATATGTTTTTGCATACAAATGTTGATCAGTTCGTAGCATTTGAATTAATACATTCTCAATCACTTTC
AAGCATTATTATGTAATAAATGATTCATGTCGAAAAGTAATAGTATCACTGTCCATTACATTTGGCAT
ATATATTTTTTTGTCAAAGCCTTACATTTGGCATATTGACGAAGCAGTTTTGTATTCACTTATATTTT
GACATCGCTTTCACAAAAATAAATAGCTATATATGATTATTATCCATTAATTGTCTCTTTTCTTTTGC
TGACACAATTGGTTGTAAATGCAATGCCAATATCCATAGCATTTGTGTGGTGAATCTTTTTCTAAGCC
TAATAGTAAATAAATCTCAATACAAGAACCCATTTACGAACAAATCAAACCAAGTTGTGATGGGTTAG
TACTTAGTAGCCCGTTTGAAATGTAGAATTTTTGATGAGATTTTACGTTTTATATAGATTTTTCTCAG
AAAACAAAAAATTCTTGCATCTTGCATTTTGGTCATTTGTAAATATTTTTTTAGTCTTAAAAAAGACC
CAAATTCTTATTAATTTCAAAATTTTCGGTCTCTAATACCTCCGGTTTTAAAAAAAAACATATCAGTT
GAAGGATGAGTTTGGTGAAGGCTATATTGTCCATTGATTTTGGAGATATATGTATTATGGTCATGATT
ATTACGATTTTTATATAAAAGAATATTAAAAATGGTGGGGTTGGTGAAGAAATGAAGATTTATCGTCA
AATATTTCAATTTTTACTTGGACTATTGCTTCGGTTATATCGTCAACATGGGCCCACTCTTCCACCAA
AGCCCAATCAATATATCTCTCGCTATCTTCACCAACCCACTCTTCTTCTCTTACCAAACCCATTTCCT
TTATTTCCAACCCTACCCCTTTATTTCTCAAGCTTTACACTTTTAGCCCATAACTTTCTTTTTATCCA
AATGGATTTGACTGGTCTCCAAAGTTGAATTAAATGGTTGTAGAAATAAAATAAAATTATACGGGTTC
AATTGTTCAATTGTTCATATACCGTTGACGTTCAATTGTTCATATACGGGTTCCGTGGTCGTTGGTAA
TATATATGTCTTTTATGCAACCAAAATAGACCAAATCAACAACAAATGAAGAAATTGTTAGAGTATGA
TACACTCATATATACCCAAATATAGCATATATTTATAATATAACTTTTGGCTATGTCATTTTACATGA
TTTTTTTGGCTTATCTATTAAAAGTATCATACAAACTGTTTTTACTTCTTTTTTTTCTTAGAATATAT
ATGCCCAAAATGGAAAAGAACATATGCCAAGGTTGATTTTATCGCTTATATGGTAAAAATTGGAAAAA
CATACAAATCATTACTTTATTTAATTAAATCATGTGAAGAAACATATTCAATTACGGTAATACGTTAT
CAAAACATTTTTTTTTACATTAATTGTTACATTTTTTTTTTTTGCAAATATTCTTAAATAACCATTCT
TTTTTTATTTACTATAATTAACATAAAAATAAATAAAATATAACATTTCAACAAAGAAATTTGCTTAT
GAAAAATACAAAATCCAGTTAATTTTTCAGAAAAATACAAATTTGCTTATAAATATATTACCACTAGT
TTATGTGATTTTAAAAGAAAGAAATGCAGCTTACCAAACGCAACGTGAAAATTTGAGAAACCCATACT

CAAAAAAGATTAAATGACAAAATCACCCTCAGCAAAATCATGAAACAACAACACTAACATTTTCACCA
ACCCCACCGTCTACTCCGGTGAATTGTCTATATGAACTCCTCCGATACAACTCCTGTTTCCTTCAGGC
CAAAGCCTAAAATTCACACAACCAAAAAAACCAACCTTTTTTTTCCACCTAAATCTTTGAATATCACA
ATATTTACTATTTACA

SEQ ID NO:28 – AtCcoAOMT promoter polynucleotide sequence

ACACATTAAAACAAAAACCATTTCCACATAAAAAAAAACGATCCAGTAAATGAAATAGATTCAAGACC
GATCGTCGAGCGGTAGAGAAAGTAAACAAAACAAAGACAGAGAATTGAAGAAACTGTGTACCTGCAAA
AATACCAATCAGATGGGTCTCCGCCAAAGTAATCTGCTTAGAAGTTTTGTAAGAAAAAACAATTAAAG
GCGTTTCATTTATTGAATTTTCCGGTTGTTTGATTCTCAGGATGAGATTGCCTATTTCCTTCAAAAAA
GAACTCTTTAATTTACACAGAAAAGCTCTGAAAATTTCCACAGAAAATGAAGAAAGAAAAGAGCGTAA
AAGGGGAAAGAGATGAAATGGGTTATTAAAAAAAGAAGCAGTGGATGAGGGAAGAGAGGATTAAGAGG
CGTAGAGATTACATGTGATGAATGATACTATCTTTTCTTACAAACACATTTTCGTGTAATTAAAATTT
AATTTGGTTCCAAAGATTTTAATCAAAAGAAGTTTGGTAAATTGAAACAGGCAGACATAATTTATTGT
AAAGAGTTTTTATTTATTTATTCATGACGTTGCTTGATGGTGCTTTACCAATTTTCTTCTCCTACGTT
AGATTTTTTTCACTTTTTTTTTTGGTGTTTGTAATAAATGTGAAAAATGGACCGTTTAAAAACTTAAA
GACGTTTGATTACTATATAAAGTAATTGTTTATAATAGAAAGTTAATTGAGACGTGAAATGGTATAAT
ATTATTGTGTAACAGTTGTGTACACGTAGCTCTCATGCAGTTTTAGTGGACCCATATGGCTTGACTTG
TATTCTGTTTTTGGGCTATTAAAGTCCAAAACAGAGACCCCTCTCAAGCCCTTCCTATTAATCCATCT
AGCTAATAGAAACTATAAACGTGTCCTCTCTCTCAATTAAATAAGCTAGAAACATACTCAACCATTCG
CATTACGCACTTCATAGCGGTAGGTTTAGATTTGTCTAAAATACTTAAAAAAATTTTTGTCTAAGTTG
TTGTCCGTTACAAAGTTTTTTTCTTTGTGACAACTTGACAACATTGACAAATAGAAAAATAAATTTCG
ATGAAACCTATGAAATGGGCTATGGCCCAACTAAAAAGAGTGGGAAATTAAAGATGGGATGGTTCAAG
TGTATACTTCGAACTTCCGACATTAGGGTCAAAGGATTTTTAAAAGGCAACCATTTGTTCCACTTTCT
CGAACAAAAACGAGCCATTTATTAATATATAGTACGGCTGAATTGGTTTTGTTCGTCATTGTGTAAAC
ACAAAGTCATTCGAATTATGTTAGGGTCCGTTGATAATATAGACGGCCCATCCCACGCACATATTAAG
TGTTCAACTCCATAGAATATCATATGGGACACTGTTTTTAATTTATAATCACCATTTAAAATGTTTAA
ATGTTTATGCAAATTGGATGGCTTCTTCACACAACATTTATTTATTGGCCTTTCATTCCATCAAAGTA
AAATAGCTTTTCAAATACATTATACTCTATACTCCTATACATGTAAATAACCATATGCATATATATTT
TTTTCAAATATAGGTCAACGCCATTTAATATAATTTTAAAAAAATTTGTTCGGAAAATATCACATTTC
TTTCACTAGACAAGCCTTGTTACCACACAATGTATCAATATGATCTAAAGGGCAAACGAAAGATCCTG
ACATGAAACGTTTAATTCTCATTTTCTCCAAATTTTATTTTTTATGTGAAGTAGATAAATTAGTATAT
ATATATATATACCAAACTAGTGTGTTATGTTATGGCAAATGTTATATCAATTCGAAGGTTCCGCTATT
GCAATATTCATTAATTTTTTCATACCAATACTATTTTTCTTTCTCTTTTATTTTGTTTTTTAATAAAT
AAAAGAAATTAAGGATGATTAGTAAGGAAGTCGCCTACCAAGAGATTCACCTACCACGGTACACTTCA
ACACCGAAGCAGAGTTGTTGAATCCACTTTTTATTCCCTTCTCTAATCTCTACTCACCAAGTCTCCAC
TTTTTTTTCTCTTTATTATATACATTTAAATTATTTAATATACGCCAACTACATACATATCCAGTGTA
ATTTCTCGTTACGTCACACCCCTTTCGTAATCGTCTAATTTCAGAAAAATATCCAGAGGTTTAAATAC
ATATTCCCATCATTAAATCTAGACATAAACACATCATACTCACAAAATTTGGCAGCAAACAGTTACTA
CAGACCCATAAATGAAAAAACGTATTCACTTGTTTTCAATTTTCACATAACCACTTCCCTGAGTTTGG
TCTCAATTTGATTGCCCCGCCGAGGCATTACTACGCCAAGTGCGATTAAGGTCCCATACAGTGTAACG
GGACCCACTATAAGACAGCGACCGACCAATTGCGTGTTAGGAGAGTTTCACCAACCCCGGACCGGTTT
TTACCGGATATAACAGAACCGGTACGAACCGGTCTCATTATCTTCCATCTTCTTTATATAGACCTCAT
GCCATGTGTGTGACTCACCAAGAAAAACACAATCGTTTAATCTCACCCAAGAAGACAAAAACACAGAG
AGAGAAAGAGAGAGAA

SEQ ID NO:29 – TcPAM amino acid sequence (*Taxus chinensis* phenylalanine aminomutase; AAT47186)

MGFAVESRSHVKDILGLINTFNEVKKITVDGTTPITVAHVAALARRHDVKVALEAEQCRARV
ETCSSWVQRKAEDGADIYGVTTGFGACSSRRTNQLSELQESLIRCLLAGVFTKGCASSVDEL
PATATRSAMLLRLNSFTYGCSGIRWEVMEALEKLLNSNVSPKVPLRGSVSASGDLIPLAYIA
GLLIGKPSVVARIGDDVEVPAPEALSRVGLRPFKLQAKEGLALVNGTSFATALASTVMYDAN
VLLLLVETLCGMFCEVIFGREEFAHPLIHKVKPHPGQIESAELLEWLLRSSPFQDLSREYYS
IDKLKKPKQDRYALRSSPQWLAPLVQTIRDATTTVETEVNSANDNPIIDHANDRALHGANFQ
GSAVGFYMDYVRIAVAGLGKLLFAQFTELMIEYYSNGLPGNLSLGPDLSVDYGLKGLDIAMA
AYSSELQYLANPVTTHVHSAEQHNQDINSLALISARKTEEALDILKLMIASHLTAMCQAVDL

RQLEEALVKVVENVVSTLADECGLPNDTKARLLYVAKAVPVYTYLESPCDPTLPLLLGLEQS
CFGSILALHKKDGIETDTLVDRLAEFEKRLSDRLENEMTAVRVLYEKKGHKTADNNDALVRI
QGSRFLPFYRFVREELDTGVMSARREQTPQEDVQKVFDAIADGRITVPLLHCLQGFLGQPNG
CANGVESFQSVWNKSA

SEQ ID NO:30 – PDC amino acid sequence (*Pediococcus pentosaceus* Phenylacrylic decarboxylase; CAC16794)
MEKTFKTLDDFLGTHFIYTYDNGWEYEWYAKNDHTVDYRIHGGMVAGRWVKDQEAHIAMLTE
GIYKVAWTEPTGTDVALDFVPNEKKLNGTIFFPKWVEEHPEITVTFQNEHIDLMEESREKYE
TYPKLVVPEFATITYMGDAGQDNDEVIAEAPYEGMTDDIRAGKYFDENYKRINK

SEQ ID NO:31 – CHS amino acid sequence (*Physcomitrella patens* chalcone synthase; ABB84527)
MASAGDVTRAALPRAQPRAEGPACVLGIGTAVPPAEFLQSEYPDFFFNITNCGEKEALKAKF
KRICDKSGIRKRHMFLTEEVLKANPGICTYMEPSLNVRHDIVVVQVPKLAAEAAQKAIKEWG
GRKSDITHIVFATTSGVNMPGADHALAKLLGLKPTVKRVMMYQTGCFGGASVLRVAKDLAEN
NKGARVLAVASEVTAVTYRAPSENHLDGLVGSALFGDGAGVYVVGSDPKPEVEKPLFEVHWA
GETILPESDGAIDGHLTEAGLIFHLMKDVPGLISKNIEKFLNEARKPVGSPAWNEMFWAVHP
GGPAILDQVEAKLKLTKDKMQGSRDILSEFGNMSSASVLFVLDQIRHRSVKMGASTLGEGSE
FGFFIGFGPGLTLEVLVLRAAPNSA

SEQ ID NO:32 – CHS amino acid sequence (*Arabidopsis thaliana* chalcone synthase; AAA32771)
MVMACASSLDEIRQAQRADGPAGILAIGTANPENHVLQAEYPDYYFRITNSEHMTDLKEKFK
RMCDKSTIRKRHMHLTEEFLKENPHMCAYMAPSLDTRQDIVVVEVPKLGKEAAVKAIKEWGQ
PKSKITHVVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRIAKDLAENN
RGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFSDGAAALIVGSDPDTSVGEKPIFEMVSA
AQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIVKSLDEAFKPLGISDWNSLFWIAHP
GGPAILDQVEIKLGLKEEKMRATRHVLSEYGNMSSACVLFILDEMRRKSAKDGVATTGEGLE
WGVLFGFGPGLTVETVVLHSVPL

SEQ ID NO:33 – SPS amino acid sequence (*Vitis vinifera* stilbene synthase; ABE68894)
MASVEEFRNAQRAKGPATILAIGTATPDHCVYQSDYADFYFRVTKSEHMTALKKKFNRICDK
SMIKKRYIHLTEEMLEEHPNIGAYMAPSLNIRQEIITAEVPKLGKEAALKALKEWGQPKSKI
THLVFCTTSGVEMPGADYKLANLLGLEPSVRRVMLYHQGCYAGGTVLRTAKDLAENNAGARV
LVVCSEITVVTFRGPSEDALDSLVGQALFGDGSAAVIVGSDPDISIERPLFQLVSAAQTFIP
NSAGAIAGNLREVGLTFHLWPNVPTLISENIEKCLTQAFDPLGISDWNSLFWIAHPGGPAIL
DAVEAKLNLDKKKLEATRHVLSEYGNMSSACVLFILDEMRKKSLKGERATTGEGLDWGVLFG
FGPGLTIETVVLHSIPMVTN

SEQ ID NO:34 – CUS amino acid sequence (*Oryza sativa* curcuminoid synthase short version; 3OIT_A)
MRRSQRADGLAAVLAIGTANPPNCVTQEEIPDFYFRVTNSDHLTALKDKFKRICQEMGVQRR
YLHHTEEMLSAHPEFVDRDAPSLDARLDIAADAVPELAAEAAKKAIAEWGRPAADITHLVVT
TNSGAHVPGVDFRLVPLLGLRPSVRRTMLHLNGCFAGCAALRLAKDLAENSRGARVLVVAAE
LTLMYFTGPDEGCFRTLLVQGLFGDGAAAVIVGADADDVERPLFEIVSAAQTIIPESDHALN
MRFTERRLDGVLGRQVPGLIGDNVERCLLDMFGPLLGGDGGGWNDLFWAVHPGSSTIMDQV
DAALGLEPGKLAASRRVLSDYGNMSGATVIFALDELRRQRKEAAAAGEWPELGVMMAFGPGM
TVDAMLLHATSHVN

SEQ ID NO:35 – CUS amino acid sequence (*Oryza sativa* curcuminoid synthase long version; 3ALE_A)

MAPTTTMGSALYPLGEMRRSQRADGLAAVLAIGTANPPNCVTQEEIPDFYFRVTNSDHLTAL
KDKFKRICQEMGVQRRYLHHTEEMLSAHPEFVDRDAPSLDARLDIAADAVPELAAEAAKKAI
AEWGRPAADITHLVVTTNSGAHVPGVDFRLVPLLGLRPSVRRTMLHLNGCFAGCAALRLAKD
LAENSRGARVLVVAAELTLMYFTGPDEGCFRTLLVQGLFGDGAAAVIVGADADDVERPLFEI
VSAAQTIIPESDHALNMRFTERRLDGVLGRQVPGLIGDNVERCLLDMFGPLLGGDGGGWND
LFWAVHPGSSTIMDQVDAALGLEPGKLAASRRVLSDYGNMSGATVIFALDELRRQRKEAAAA
GEWPELGVMMAFGPGMTVDAMLLHATSHVN

SEQ ID NO:36 – BAS amino acid sequence (*Rheum palmatum* benzalacetone synthase; AAK82824)

MATEEMKKLATVMAIGTANPPNCYYQADFPDFYFRVTNSDHLINLKQKFKRLCENSRIEKRY
LHVTEEILKENPNIAAYEATSLNVRHKMQVKGVAELGKEAALKAIKEWGQPKSKITHLIVCC
LAGVDMPGADYQLTKLLDLDPSVKRFMFYHLGCYAGGTVLRLAKDIAENNKGARVLIVCSEM
TTTCFRGPSETHLDSMIGQAILGDGAAAVIVGADPDLTVERPIFELVSTAQTIVPESHGAIE
GHLLESGLSFHLYKTVPTLISNNIKTCLSDAFTPLNISDWNSLFWIAHPGGPAILDQVTAKV
GLEKEKLKVTRQVLKDYGNMSSATVFFIMDEMRKKSLENGQATTGEGLEWGVLFGFGPGITV
ETVVLRSVPVIS

SEQ ID NO:37 – AtPAP1 amino acid sequence (*Arabidopsis thaliana* R2R3 Myb transcription factor, AtMyb75; AAG42001)

MEGSSKGLRKGAWTTEEDSLLRQCINKYGEGKWHQVPVRAGLNRCRKSCRLRWLNYLKPSIK
RGKLSSDEVDLLLRLHRLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHEPCCKIKMKKRD
ITPIPTTPALKNNVYKPRPRSFTVNNDCNHLNAPPKVDVNPPCLGLNINNVCDNSIIYNKDK
KKDQLVNNLIDGDNMWLEKFLEESQEVDILVPEATTTEKGDTLAFDVDQLWSLFDGETVKFD

SEQ ID NO:38 – AtPAP2 amino acid sequence (*Arabidopsis thaliana* R2R3 Myb transcription factor, AtMyb90; AAG42002)

MEGSSKGLRKGAWTAEEDSLLRLCIDKYGEGKWHQVPLRAGLNRCRKSCRLRWLNYLKPSIK
RGRLSNDEVDLLLRLHKLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHESSCCKSKMKKK
NIISPPTTPVQKIGVFKPRPRSFSVNNGCSHLNGLPEVDLIPSCLGLKKNNVCENSITCNKD
DEKDDFVNNLMNGDNMWLENLLGENQEADAIVPEATTAEHGATLAFDVEQLWSLFDGETVEL
D

SEQ ID NO:39 – AtTT2 amino acid sequence (*Arabidopsis thaliana* R2R3 Myb transcription factor, AtMyb123; AED93980)

MGKRATTSVRREELNRGAWTDHEDKILRDYITTHGEGKWSTLPNQAGLKRCGKSCRLRWKNY
LRPGIKRGNISSDEEELIIRLHNLLGNRWSLIAGRLPGRTDNEIKNHWNSNLRKRLPKTQTK
QPKRIKHSTNNENNVCVIRTKAIRCSKTLLFSDLSLQKKSSTSPLPLKEQEMDQGGSSLMGD
LEFDFDRIHSEFHFPDLMDFDGLDCGNVTSLVSSNEILGELVPAQGNLDLNRPFTSCHHRGD
DEDWLRDFTC

SEQ ID NO:40 – NtAn2 amino acid sequence (*Nicotiana tabacum* R2R3 Myb transcription factor; ACO52470)

MNICTNKSSSGVKKGAWTEEEDVLLKKCIEKYGEGKWHQVPLRAGLNRCRKSCRLRWLNYLR
PHIKRGDFSFDEVDLILRLHKLLGNRWSLIAGRLPGRTANDVKNYWNSHLRKKLIAPHDQKE
SKQKAKKITIFRPRPRTFSKTNTCVKSNTNTVDKDIEGSSEIIRFNDNLKPTTEELTDDGIQ
WWADLLANNYNNNGIEEADNSSPTLLHEEMPLLS

SEQ ID NO:41 – MtLAP1 amino acid sequence (*Medicago truncatula* R2R3 Myb transcription factor; ACN79541)

MENTGGVRKGAWTYKEDELLKACINTYGEGKWNLVPQRSGLNRCRKSCRLRWLNYLSPNINR
GRFSEDEEDLILRLHKLLGNRWSLIAGRLPGRTANDVKNYWHTNLAKKVVSEKEEEKENDKP

KETMKAHEVIKPRPITLSSHSNWLKGKNSIPRDLDYSENMASNQIGRECASTSKPDLGNAPI
PCEMWCDSLWNLGEHVDSEKIGSCSSLQEENLMEFPNVDDDSFWDFNLCDLNSLWDLP

SEQ ID NO:42 – ZmMYB-C amino acid sequence (*Zea mays* R2R3 Myb transcription factor; AAK09326)
MGRRACCAKEGVKRGAWTSKEDDALAAYVKAHGEGKWREVPQKAGLRRCGKSCRLRWLNYLR
PNIRRGNISYDEEDLIIRLHRLLGNRWSLIAGRLPGRTDNEIKNYWNSTLGRRAGAGAGAGG
SWVVVAPDTGSHATPAATSGACETGQNSAAHRADPDSAGTTTTSAAAVWAPKAVRCTGGLFF
FHRDTTPAHAGETATPMAGGGGGGGGEAGSSDDCSSAASVSLRVGSHDEPCFSGDGDGDWMD
DVRALASFLESDEDWLRCQTAGQLA

SEQ ID NO:43 – ZmMYC-Lc amino acid sequence (*Zea mays* BHLH transcription factor; ABD72707)
MALSASRVQQAEELLQRPAERQLMRSQLAAAARSINWSYALFWSISDTQPGVLTWTDGFYNG
EVKTRKISNSVELTSDQLVMQRSDQLRELYEALLSGEGDRRAAPARPAGSLSPEDLGDTEWY
YVVSMTYAFRPGQGLPGRSFASDEHVWLCNAHLAGSKAFPRALLAKSASIQSILCIPVMGGV
LELGTTDTVPEAPDLVSRATAAFWEPQCPSSSPSGRANETGEAAADDGTFAFEELDHNNGMD
DIEAMTAAGGHGQEEELRLREAEALSDDASLEHITKEIEEFYSLCDEMDLQALPLPLEDGWT
VDASNFEVPCSSPQPAPPPVDRATANVAADASRAPVYGSRATSFMAWTRSSQQSSCSDDAAP
AAVVPAIEEPQRLLKKVVAGGGAWESCGGATGAAQEMSGTGTKNHVMSERKRREKLNEMFLV
LKSLLPSIHRVNKASILAETIAYLKELQRRVQELESSREPASRPSETTTRLITRPSRGNNES
VRKEVCAGSKRKSPELGRDDVERPPVLTMDAGTSNVTVTVSDKDVLLEVQCRWEELLMTRVF
DAIKSLHLDVLSVQASAPDGFMGLKIRAQFAGSGAVVPWMISEALRKAIGKR

SEQ ID NO:44 – AtTT8 amino acid sequence (*Arabidopsis thaliana* BHLH transcription factor; AEE82802)
MDESSIIPAEKVAGAEKKELQGLLKTAVQSVDWTYSVFWQFCPQQRVLVWGNGYYNGAIKTR
KTTQPAEVTAEEAALERSQQLRELYETLLAGESTSEARACTALSPEDLTETEWFYLMCVSFS
FPPPSGMPGKAYARRKHVWLSGANEVDSKTFSRAILAKSAKIQTVVCIPMLDGVVELGTTKK
VREDVEFVELTKSFFYDHCKTNPKPALSEHSTYEVHEEAEDEEEVEEEMTMSEEMRLGSPDD
EDVSNQNLHSDLHIESTHTLDTHMDMMNLMEEGGNYSQTVTTLLMSHPTSLLSDSVSTSSYI
QSSFATWRVENGKEHQQVKTAPSSQWVLKQMIFRVPFLHDNTKDKRLPREDLSHVVAERRRR
EKLNEKFITLRSMVPFVTKMDKVSILGDTIAYVNHLRKRVHELENTHHEQQHKRTRTCKRKT
SEEVEVSIIENDVLLEMRCEYRDGLLLDILQVLHELGIETTAVHTSVNDHDFEAEIRAKVRG
KKASIAEVKRAIHQVIIHDTNL

SEQ ID NO:45 – VvMyc1 amino acid sequence (*Vitis vinifera* BHLH transcription factor; ACC68685)
MAAPPNSRLQSMLQSAVQSVRWTYSLFWQICPQQGILVWGDGYYNGAIKTRKTVQPMEVSAE
EASLQRSQQLRELYESLSAGETNQPARRPCAALSPEDLTESEWFYLMCVSFSFPPGVGLPGK
AYAKRHHIWLAGANEVDSKVFSRAILAKSARVQTVVCIPLMDGVVEFGTTEKVQEDLGFVQH
VKSFFTDHHLHNHPPKPALSEHSTSNPATSSDHSRFHSPPIQAAYAAADPPASNNQEEEEEE
EEEEEEEEEEEEEEEEEEEAESDSEAETGRNNRRVRTQNTGTEGVAGSHTAAEPSELIQLEMS
EGIRLGSPDDGSNNLDSDFHMLAVSQPGSSVDHQRRADSYRAESARRWPMLQDPLCSSGLQQ
PPPQPPTGPPPLDELSQEDTHYSQTVSTILQHQPNRWSESSSSGCIAPYSSQSAFAKWTTRC
DHHHHPMAVEGTSQWLLKYILFSVPFLHTKYRDENSPKSRDGDSAGRFRKGTPQDELSANHV
LAERRRREKLNERFIILRSLVPFVTKMDKASILGDTIEYVKQLRKKIQDLEARTRQMEVEQR
SRGSDSVRSKEHRIGSGSVDRNRAVVAGSDKRKLRIVEGSTGAKPKVVDSPPAAVEGGTTTV
EVSIIESDALLEMQCPYREGLLLDVMQMLRELRLETTTVQSSLTNGVFVAELRAKVKENASG
KKASIMEVKRAINQIIPQC

[0184] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the purview of this application and scope of the appended claims.

[0185] This application contains a sequence listing in electronic form in ASCII text format. A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

WHAT IS CLAIMED IS:

1. A method of engineering a plant having reduced lignin content, the method comprising:

introducing into the plant an expression cassette comprising a polynucleotide that encodes an enzyme that diverts a monolignol precursor from the lignin biosynthesis pathway, wherein the polynucleotide is operably linked to a heterologous secondary cell wall-specific promoter or a heterologous fiber cell-specific promoter, and wherein the enzyme is: a bacterial dehydroshikimate dehydratase, a *Podospora anserina* dehydroshikimate dehydratase (DsDH), a bacterial shikimate kinase, a pentafunctional AROM polypeptide (ARO1), a phenylacetaldehyde synthase (PAAS), a phenylalanine aminomutase (PAM), a *p*-coumarate/cinnamate carboxylmethlytransferase (CCMT1), a phenylacrylic acid decarboxylase (PDC), a 2-oxoglutarate-dependent dioxygenase (C2'H), a chalcone synthase (CHS), a stilbene synthase (SPS), a cucuminoid synthase (CUS), or a benzalacetone synthase (BAS); and

culturing the plant under conditions in which the enzyme that diverts the monolignol precursor from the lignin biosynthesis pathway is expressed, thereby reducing the plant lignin content.

2. The method of claim 1, wherein the enzyme reduces the amount of cytosolic and/or plastidial shikimate that is available for the lignin biosynthesis pathway.

3. The method of claim 2, wherein the polynucleotide encodes *Corynebacterium glutamicum* pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), or dehydroshikimate dehydratase (QsuB).

4. The method of claim 2, wherein the polynucleotide encodes a shikimate kinase (AroK).

5. The method of claim 2, wherein the enzyme:

has shikimate kinase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:2,

has pentafunctional AROM enzyme activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:4, or

Date Reçue/Date Received 2021-05-05

has dehydryoshikimate dehydratase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:6 or SEQ ID NO:8.

6. The method of claim 5, wherein the enzyme comprises the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

7. The method of claim 1, wherein the enzyme reduces the amount of cytosolic and/or plastidial phenylalanine that is available for the lignin biosynthesis pathway.

8. The method of claim 7, wherein the enzyme is phenylacetaldehyde synthase (PAAS) or phenylalanine aminomutase (PAM).

9. The method of claim 7, wherein the enzyme has phenylacetaldehyde synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:10; or has phenylalanine aminomutase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:29.

10. The method of claim 1, wherein the enzyme reduces the amount of cinnamate and/or coumarate that is available for the lignin biosynthesis pathway.

11. The method of claim 10, wherein the enzyme is *p*-coumarate/cinnamate carboxylmethltransferase (CCMT1) or phenylacrylic acid decarboxylase (PDC).

12. The method of claim 10, wherein the enzyme has *p*-coumarate/cinnamate carboxylmethyltransferase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:12, or has phenylacrylic acid decarboxylase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:30.

13. The method of claim 1, wherein the enzyme reduces the amount of coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the lignin biosynthesis pathway.

Date Reçue/Date Received 2021-05-05

14. The method of claim 13, wherein the enzyme is 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone synthase (BAS).

15. The method of claim 13, wherein the enzyme:

has 2-oxoglutarase-dependent dioxygenase activity and comprises and amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:14,

has chalcone synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:31 or SEQ ID NO:32,

has stilbene synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:33,

has cucuminoid synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:34 or SEQ ID NO;35, or

has benzalacetone synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:36.

16. The method of claim 1, wherein the metabolic pathway is a stilbene biosynthesis pathway, a flavonoid biosynthesis pathway, a curcuminoid biosynthesis pathway, or a benzalacetone biosynthesis pathway.

17. The method of any one of claims 1 to 16, wherein the promoter is the IRX5 promoter.

18. The method of any one of claims 1 to 16, wherein the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter.

19. The method of claim 18, wherein the promoter has promoter activity and comprises a nucleic acid sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or 28.

20. The method of any of claims 1 to 19, wherein the enzyme that diverts a monolignol precursor from a lignin biosynthesis pathway is targeted to a plastid in the plant.

Date Reçue/Date Received 2021-05-05

21. The method of claim 20, wherein the polynucleotide comprises a sequence encoding a plastid targeting signal that is at least 95% identical to the polynucleotide sequence of SEQ ID NO:15.

22. The method of any of claims 1 to 21, wherein the plant is selected from the group consisting of Arabidopsis, poplar, eucalyptus, rice, corn, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, and Brachypodium.

23. A plant cell engineered by the method of any one of claims 1 to 22, wherein the plant cell comprises the expression cassette.

24. A plant cell comprising an expression cassette comprising a polynucleotide that encodes an enzyme that diverts a monolignol precursor from the lignin biosynthesis pathway, wherein the polynucleotide is operably linked to a heterologous secondary cell wall-specific promoter or a heterologous fiber cell-specific promoter, and wherein the enzyme is a bacterial dehydroshikimate dehydratase, a *Podospora anserina* dehydroshikimate dehydratase (DsDH), a bacterial shikimate kinase, a pentafunctional AROM polypeptide (ARO1), a phenylacetaldehyde synthase (PAAS), a phenylalanine aminomutase (PAM), a *p*-coumarate/cinnamate carboxylmethyltransferase (CCMT1), a phenylacrylic acid decarboxylase (PDC), a 2-oxoglutarate-dependent dioxygenase (C2'H), a chalcone synthase (CHS), a stilbene synthase (SPS), a cucuminoid synthase (CUS), or a benzalacetone synthase (BAS).

25. The plant cell of claim 24, wherein the enzyme reduces the amount of cytosolic and/or plastidial shikimate that is available for the lignin biosynthesis pathway.

26. The plant cell of claim 25, wherein the polynucleotide encodes *Corynebacterium glutamicum* pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), or dehydroshikimate dehydratase (QsuB).

27. The plant cell of claim 25, wherein the polynucleotide encodes a shikimate kinase (AroK).

Date Reçue/Date Received 2021-05-05

28. The plant cell of claim 25, wherein the enzyme:

has shikimate kinase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:2,

has pentafunctional AROM enzyme activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:4, or

has dehydryoshikimate dehydratase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:6 or SEQ ID NO:8.

29. The plant cell of claim 28, wherein the enzyme comprises the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, or SEQ ID NO:8.

30. The plant cell of claim 24, wherein the enzyme reduces the amount of cytosolic and/or plastidial phenylalanine that is available for the lignin biosynthesis pathway.

31. The plant cell of claim 30, wherein the enzyme is phenylacetaldehyde synthase (PAAS) or phenylalanine aminomutase (PAM).

32. The plant cell of claim 30, wherein the enzyme has phenylacetaldehyde synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:10; or has phenylalanine aminomutase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:29.

33. The plant cell of claim 24, wherein the enzyme reduces the amount of cinnamate and/or coumarate that is available for the lignin biosynthesis pathway.

34. The plant cell of claim 33, wherein the enzyme is *p*-coumarate/cinnamate carboxylmethltransferase (CCMT1) or phenylacrylic acid decarboxylase (PDC).

35. The plant cell of claim 33, wherein the enzyme has *p*-coumarate/cinnamate carboxylmethyltransferase activity and comprises an amino acid

Date Reçue/Date Received 2021-05-05

sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:12, or has phenylacrylic acid decarboxylase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:30.

36. The plant cell of claim 24, wherein the enzyme reduces the amount of coumaroyl-CoA, caffeoyl-CoA, and/or feruloyl-CoA that is available for the lignin biosynthesis pathway.

37. The method of claim 36, wherein the enzyme is 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone synthase (BAS).

38. The method of claim 36, wherein the enzyme:

has 2-oxoglutarase-dependent dioxygenase activity and comprises and amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:14,

has chalcone synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:31 or SEQ ID NO:32,

has stilbene synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:33,

has cucuminoid synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:34 or SEQ ID NO;35, or

has benzalacetone synthase activity and comprises an amino acid sequence having at least 95% identity to the amino acid sequence of SEQ ID NO:36.

39. The method of claim 24, wherein the metabolic pathway is a stilbene biosynthesis pathway, a flavonoid biosynthesis pathway, a curcuminoid biosynthesis pathway, or a benzalacetone biosynthesis pathway.

40. The plant cell of any one of claims 24 to 39, wherein the promoter is the IRX5 promoter.

41. The plant cell of any one of claims 24 to 39, wherein the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter.

Date Reçue/Date Received 2021-05-05

42. The plant cell of claim 41, wherein the promoter has promoter activity and comprises a nucleic acid sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, or 28.

43. The plant cell of any of claims 24 to 42, wherein the enzyme that diverts a monolignol precursor from a lignin biosynthesis pathway is targeted to a plastid in the plant.

44. The plant cell of claim 43, wherein the polynucleotide comprises a sequence encoding a plastid targeting signal that that is at least 95% identical to the polynucleotide sequence of SEQ ID NO:15.

45. The plant cell of any of claims 24 to 44, wherein the plant is selected from the group consisting of Arabidopsis, poplar, eucalyptus, rice, corn, switchgrass, sorghum, millet, miscanthus, sugarcane, pine, alfalfa, wheat, soy, barley, turfgrass, tobacco, hemp, bamboo, rape, sunflower, willow, and Brachypodium.

46. A plant cell comprising a polynucleotide that encodes an enzyme that diverts a monolignol precursor from a lignin biosynthesis pathway in a plant, wherein: the polynucleotide is operably linked to a heterologous promoter; the plant cell is a secondary cell wall or fiber cell; the enzyme is shikimate kinase (AroK), pentafunctional AROM polypeptide (ARO1), dehydroshikimate dehydratase (DsDH), dehydroshikimate dehydratase (QsuB), phenylacetaldehyde synthase (PAAS), phenylalanine aminomutase (PAM), *p*-coumarate/cinnamate carboxylmethltransferase (CCMT1), ferulic acid decarboxylase (FDC1), phenylacrylic decarboxylase (PDC), 2-oxoglutarate-dependent dioxygenase (C2'H), chalcone synthase (CHS), stilbene synthase (SPS), cucuminoid synthase (CUS), or benzalacetone synthase (BAS); and the heterologous promoter is a secondary cell wall-specific promoter or a fiber cell-specific promoter.

47. The plant cell of claim 46, wherein the enzyme is comprises an amino acid sequence having at least 95% identity to the amino acid sequence of any of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 29, 30, 31, 32, 33, 34, 35, or 36; and has shikimate kinase, pentafunctional AROM polypeptide, dehydroshikimate dehydratase, phenylacetaldehyde synthase, penylalanine aminomutase, *p*-coumarate/cinnamate carboxylmethltransferase,

Date Reçue/Date Received 2021-05-05

phenylacrylic decarboxylase, 2-oxoglutarate-dependent dioxygenase, chalcone synthase, stilbene synthase, cucuminoid synthase, or benzalacetone synthase activity.

48. The plant cell of claim 46 or 47, wherein the promoter is an IRX5 promoter.

49. The plant cell of claim 46 or 47, wherein the promoter is a C4H, C3H, HCT, CCR1, CAD4, CAD5, F5H, PAL1, PAL2, 4CL1, or CCoAMT promoter.

50. A method of obtaining an increased amount of soluble sugars from a plant in a saccharification reaction, the method comprising:

subjecting a plant generated by the method of any one of claims 1 to 22 or a plant comprising the plant cell of any one of claims 23 to 49 to a saccharification reaction, thereby increasing the amount of soluble sugars that can be obtained from the plant as compared to a wild-type plant.

51. A method of improving digestibility in ruminants, the method comprising providing a plant generated by the method of any one of claims 1 to 22 or a plant comprising the plant cell of any one of claims 23 to 49 to the ruminant.

Date Reçue/Date Received 2021-05-05

1/28

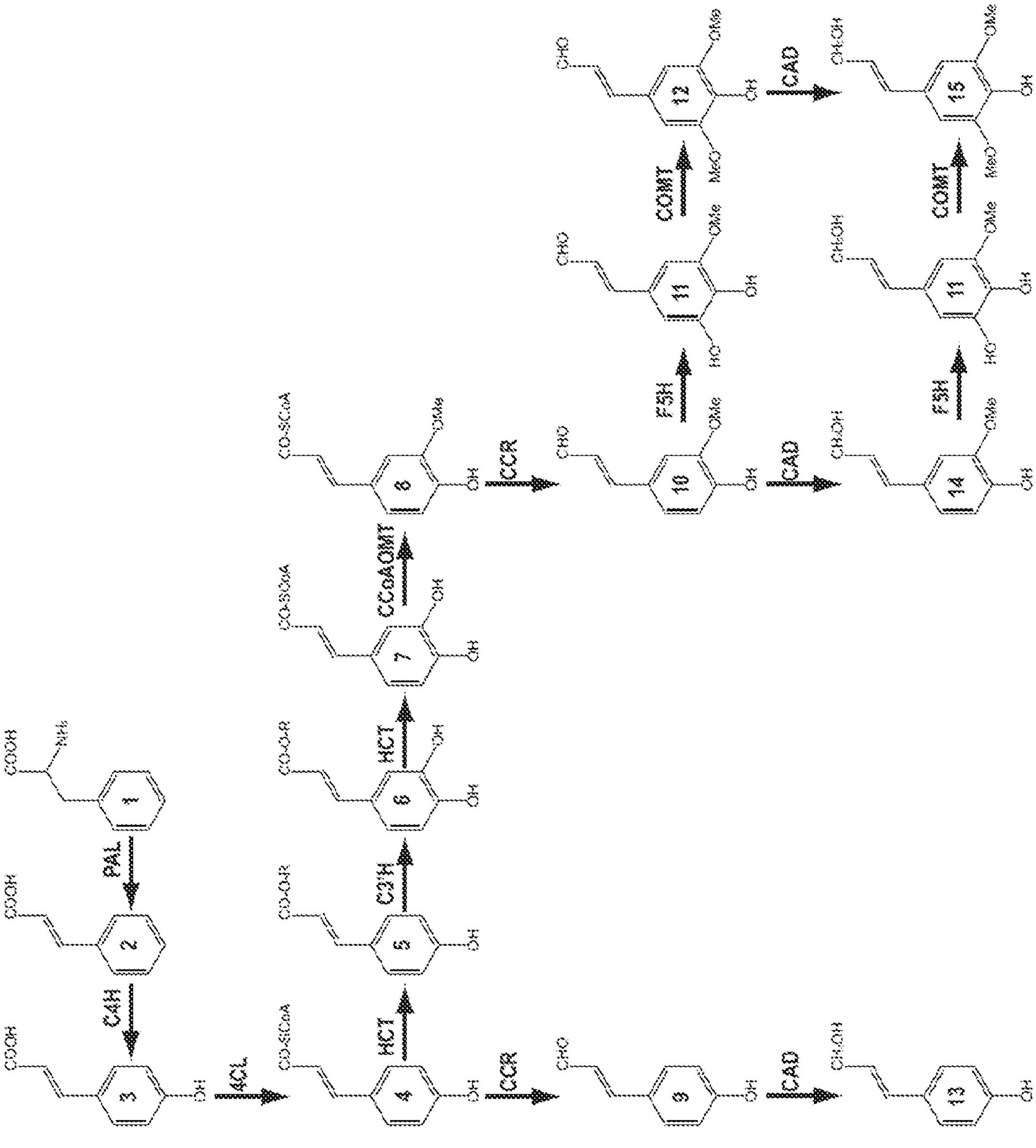

PAL
C4H
4CL
HCT
C3'H
HCT
CCoAOMT
CCR
CCR
CAD
CAD
F5H
F5H
COMT
COMT
CAD
CAD

FIG. 1

Precursor depletion strategy to repress enzymatic steps:

Depletion of shikimate: HCT co-substrate

(1) Cytosolic depletion of shikimate

MtAroK: *Mycobacterium tuberculosis* shikimate kinase

(2) Plastidial depletion of shikimate

ScAro1 + plastid targeting signal. ScAro1 (YDR17W): *Saccharomyces cerevisiae* Pentafunctional arom protein, catalyzes steps 2 through 6 in the biosynthesis of chorismate, which is a precursor to aromatic amino acids

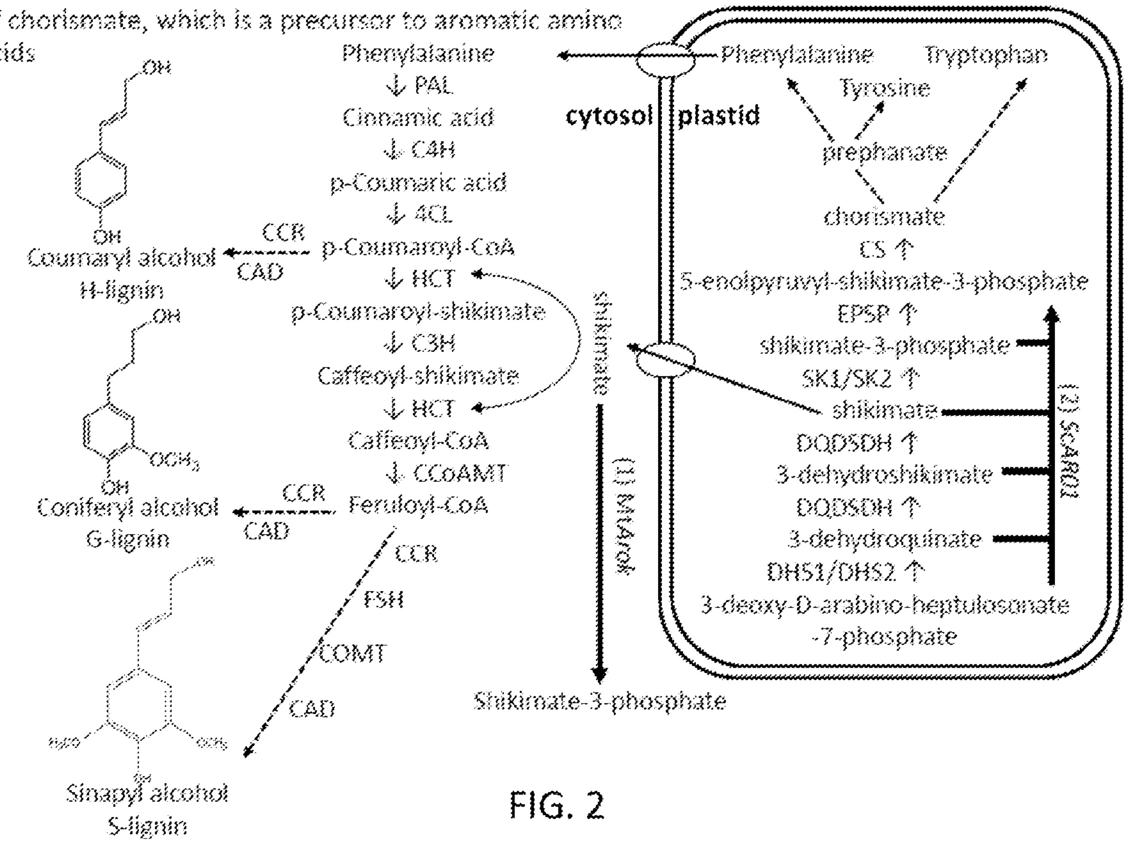


FIG. 2

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:

Depletion of shikimate and production of new stoppers

Plastidial depletion of shikimate

CgQsuB + plastid targeting signal. QsuB: *Corynebacterium glutamicum* dehydroshikimate dehydratase

PaDsDH + plastid targeting signal. DsDH: *Podospora anserina* dehydroshikimate dehydratase

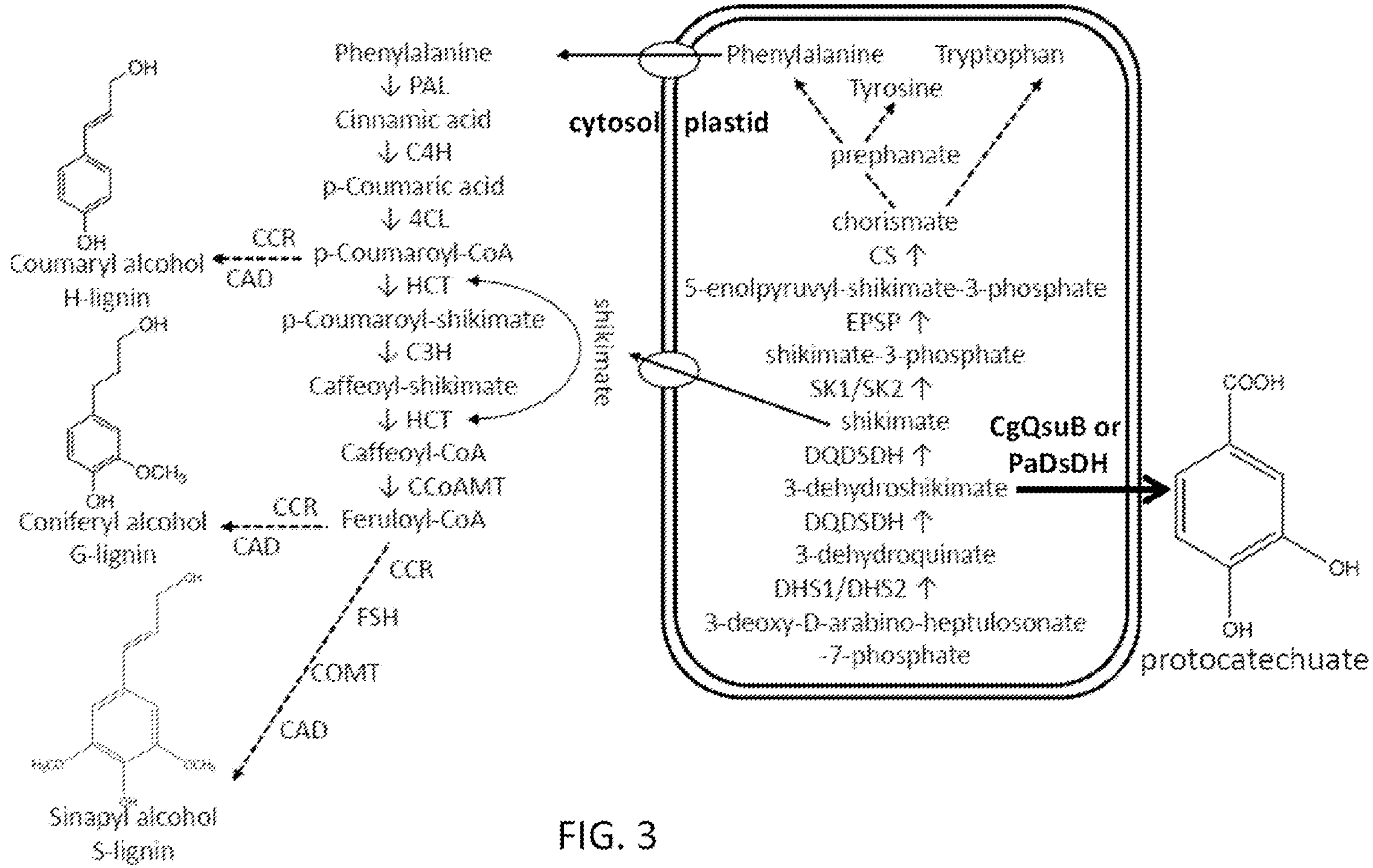


FIG. 3

4/28

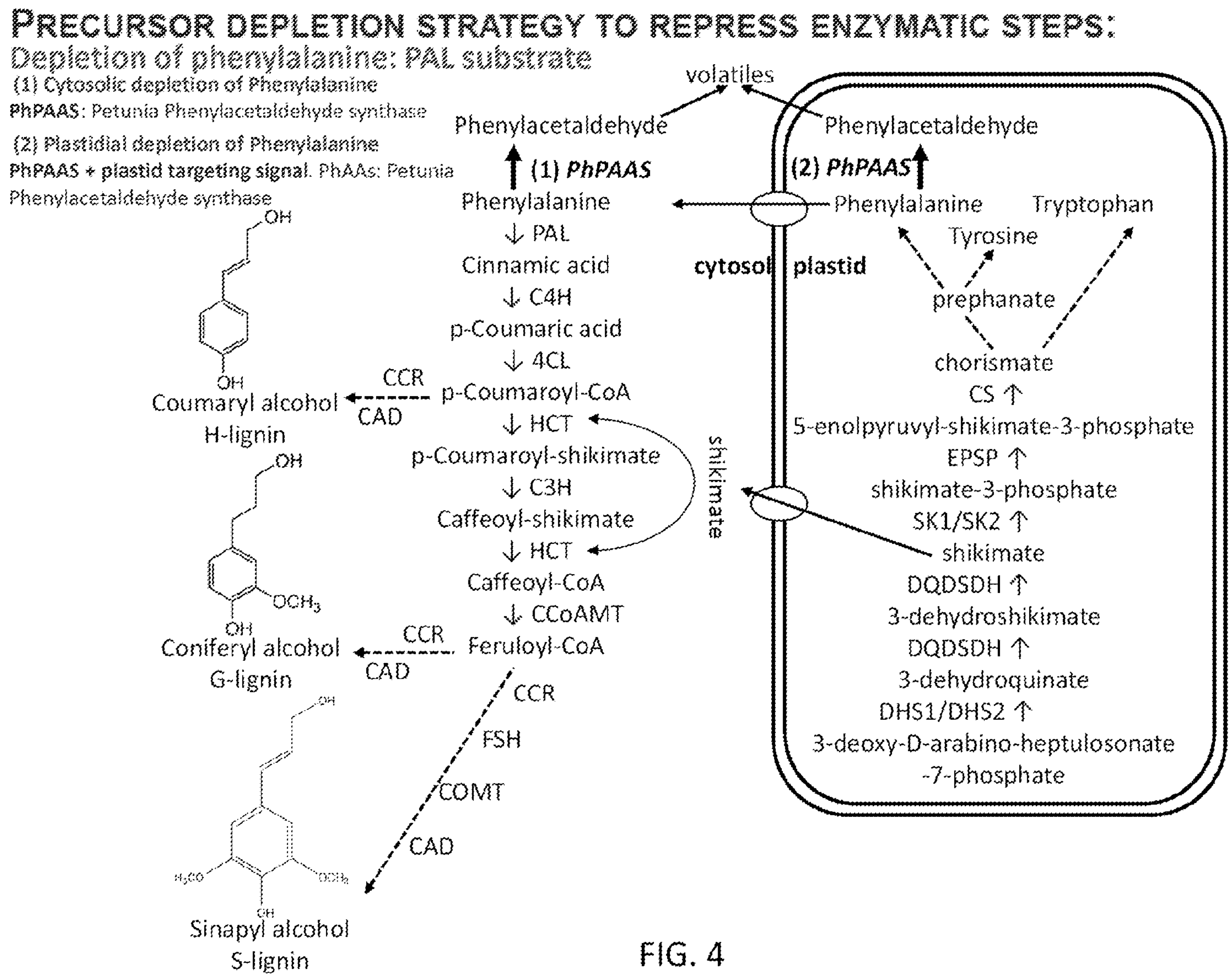

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of phenylalanine: PAL substrate
(1) Cytosolic depletion of Phenylalanine
PhPAAS: Petunia Phenylacetaldehyde synthase
(2) Plastidial depletion of Phenylalanine
PhPAAS + plastid targeting signal. PhAAs: Petunia Phenylacetaldehyde synthase
volatiles
Phenylacetaldehyde
(1) PhPAAS
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
CAD
Coumaryl alcohol
H-lignin
CCR
CAD
Coniferyl alcohol
G-lignin
CCR
FSH
COMT
CAD
Sinapyl alcohol
S-lignin
cytosol
plastid
shikimate
Phenylacetaldehyde
(2) PhPAAS
Phenylalanine
Tryptophan
Tyrosine
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate -7-phosphate

FIG. 4

5/28

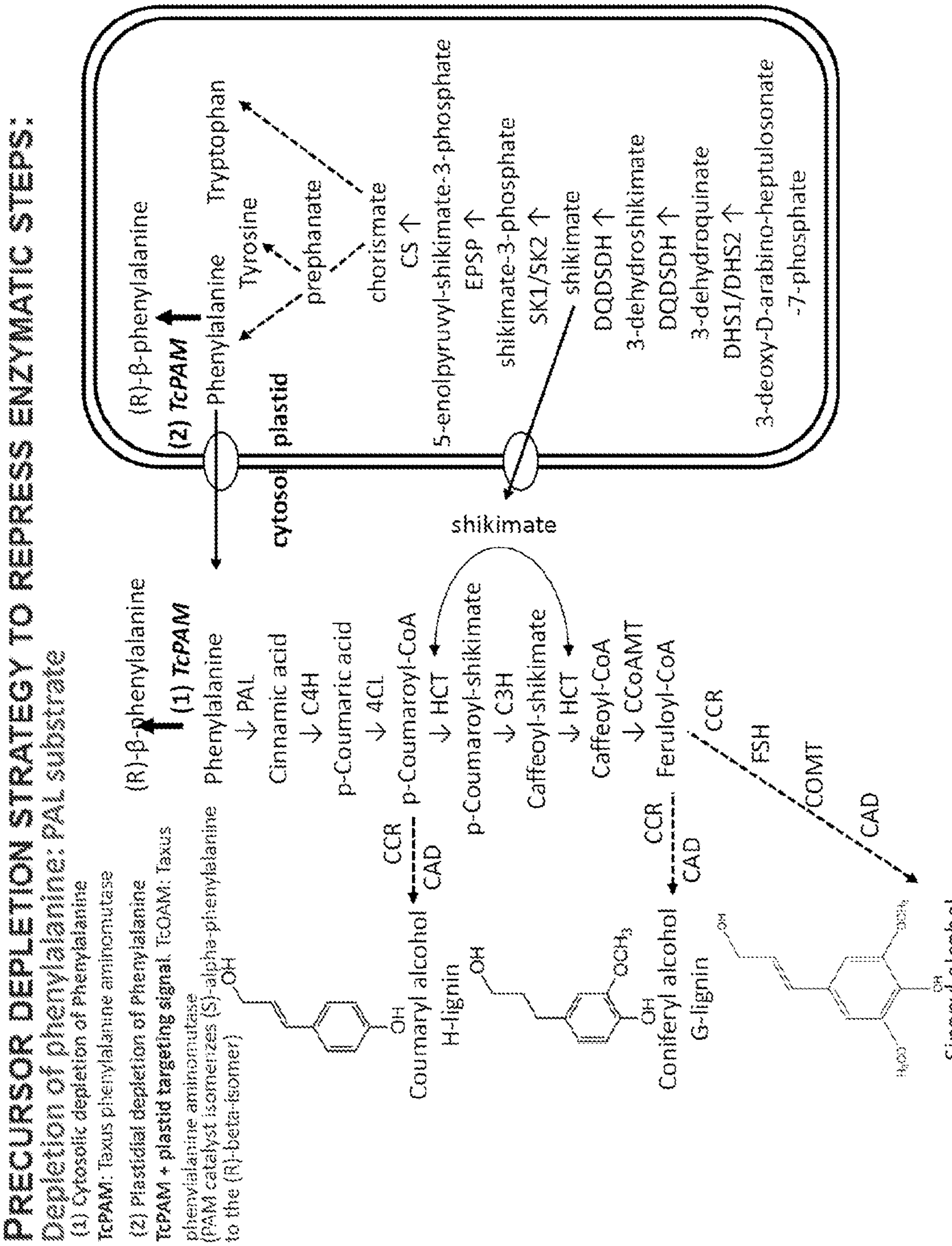

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of phenylalanine: PAL substrate
(1) Cytosolic depletion of Phenylalanine
TcPAM: Taxus phenylalanine aminomutase
(2) Plastidial depletion of Phenylalanine
TcPAM + plastid targeting signal. TcOAM: Taxus phenylalanine aminomutase (PAM catalyst isomerizes (S)-alpha-phenylalanine to the (R)-beta-isomer)
(R)-β-phenylalanine
(1) TcPAM
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
FSH
COMT
CAD
CCR
CAD
Coumaryl alcohol
H-lignin
Coniferyl alcohol
G-lignin
Sinapyl alcohol
S-lignin
cytosol
plastid
shikimate
(R)-β-phenylalanine
(2) TcPAM
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate -7-phosphate

FIG. 5

6/28

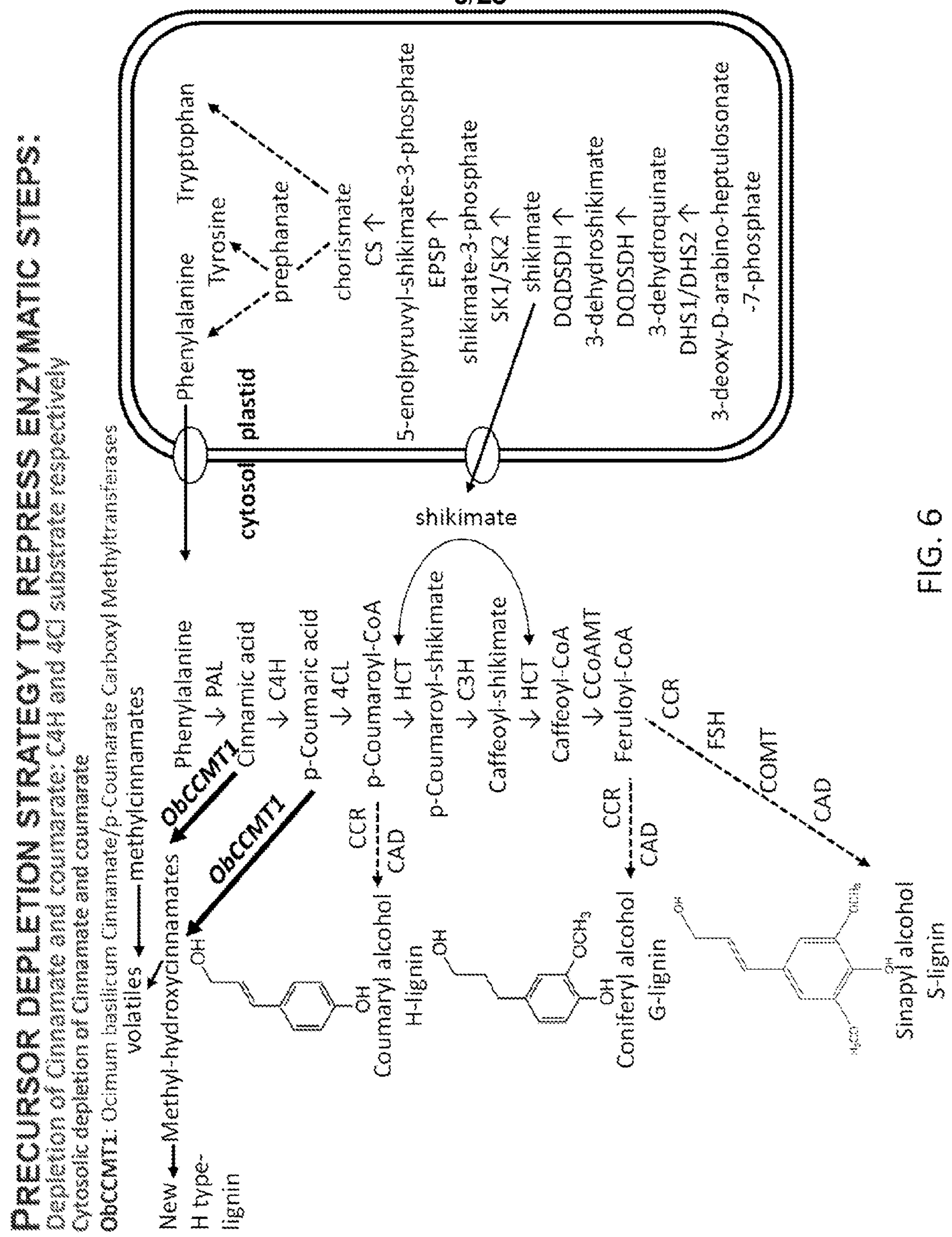

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of Cinnamate and coumarate: C4H and 4Cl substrate respectively
Cytosolic depletion of Cinnamate and coumarate
ObCCMT1: Ocimum basilicum Cinnamate/p-Coumarate Carboxyl Methyltransferases
New H type-lignin
Methyl-hydroxycinnamates
volatiles
methylcinnamates
ObCCMT1
ObCCMT1
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
FSH
COMT
CAD
CCR
CAD
CCR
CAD
Coumaryl alcohol
H-lignin
Coniferyl alcohol
G-lignin
Sinapyl alcohol
S-lignin
OH
OCH_3
cytosol
plastid
shikimate
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate -7-phosphate

FIG. 6

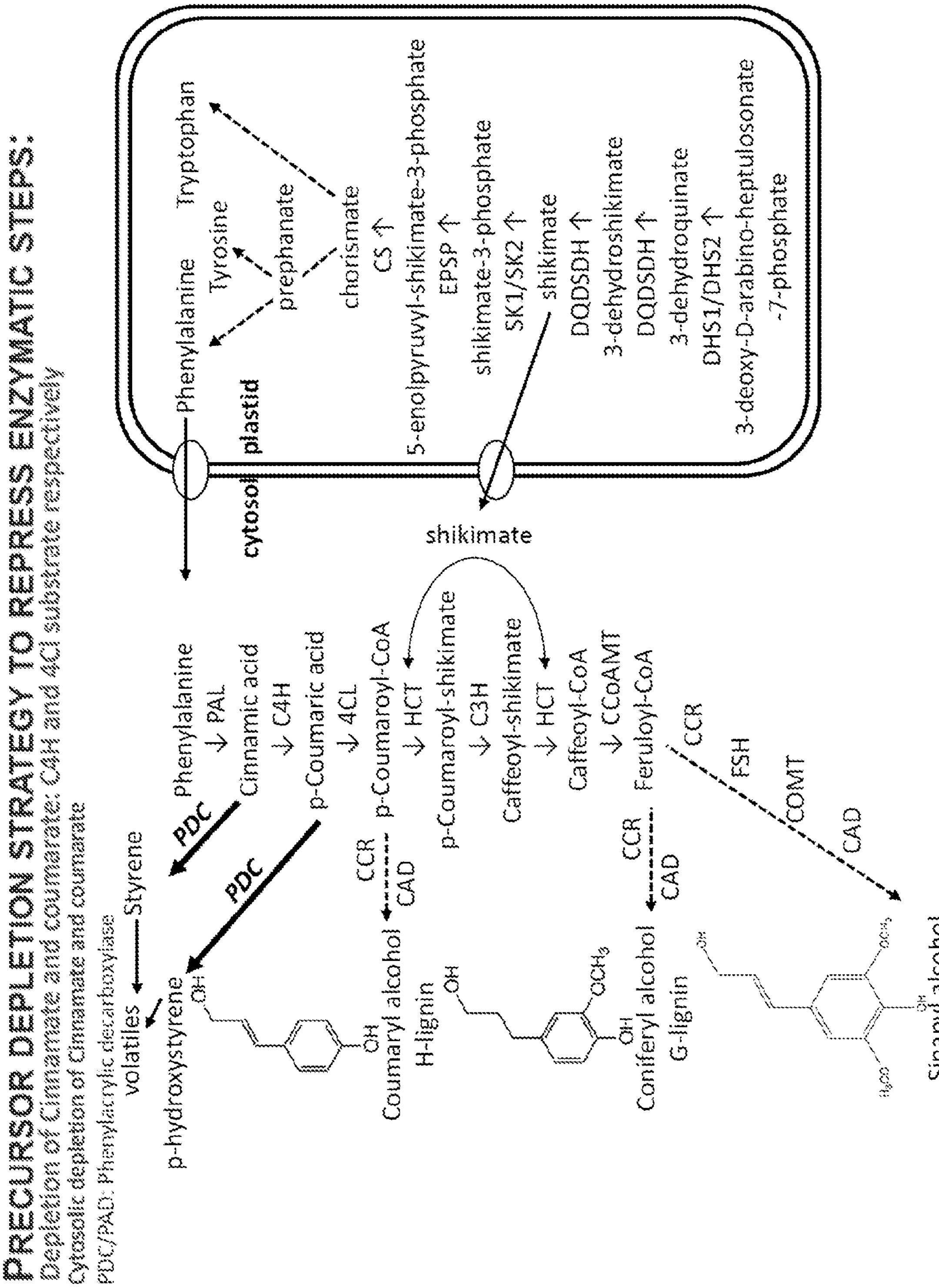

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of Cinnamate and coumarate: C4H and 4Cl substrate respectively
Cytosolic depletion of Cinnamate and coumarate
PDC/PAD: Phenylacrylic decarboxylase
cytosol
plastid
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate
-7-phosphate
shikimate
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
FSH
COMT
CAD
PDC
PDC
Styrene
volatiles
p-hydroxystyrene
CCR
CAD
Coumaryl alcohol
H-lignin
CCR
CAD
Coniferyl alcohol
G-lignin
Sinapyl alcohol
S-lignin

FIG. 7

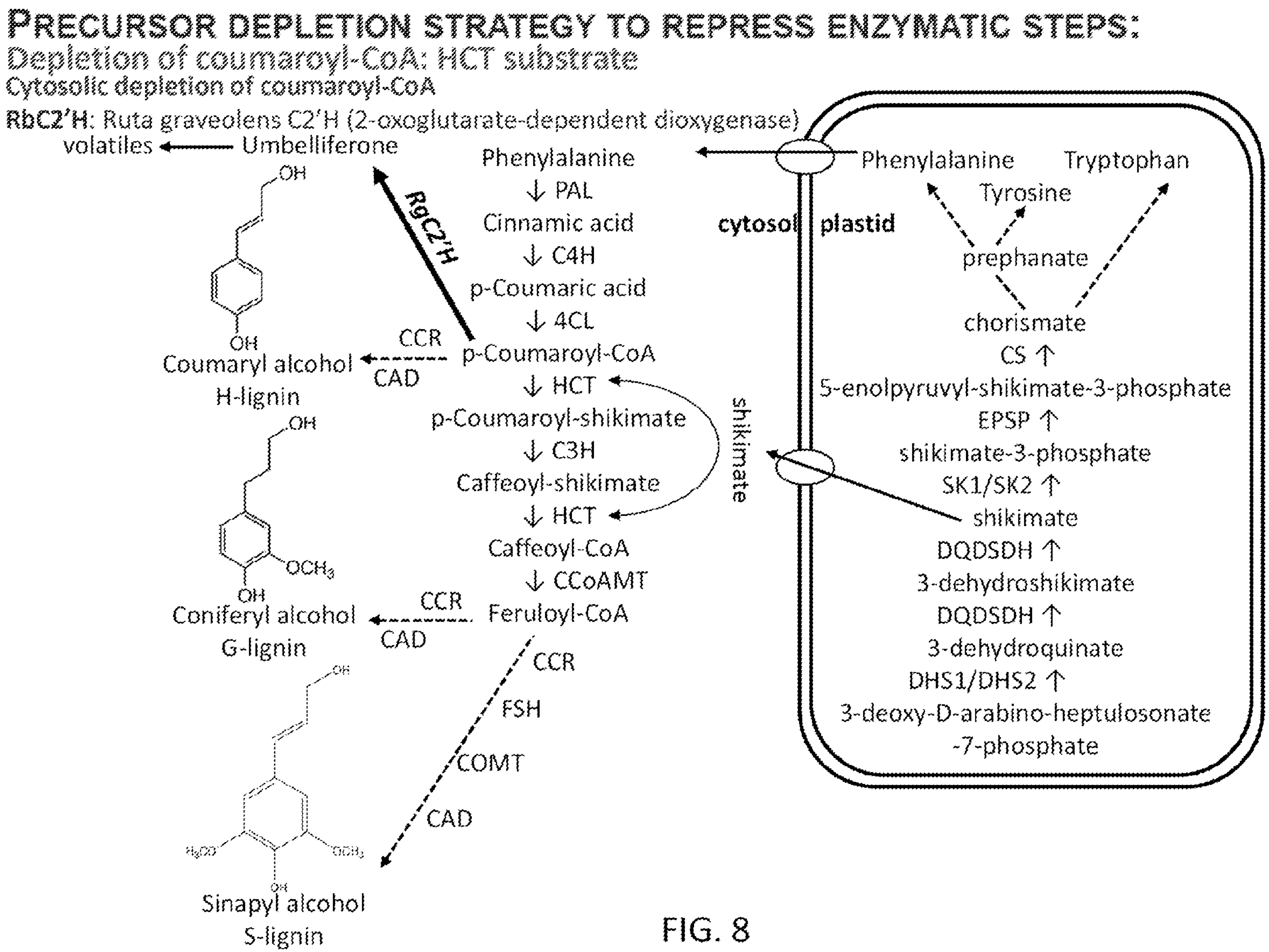
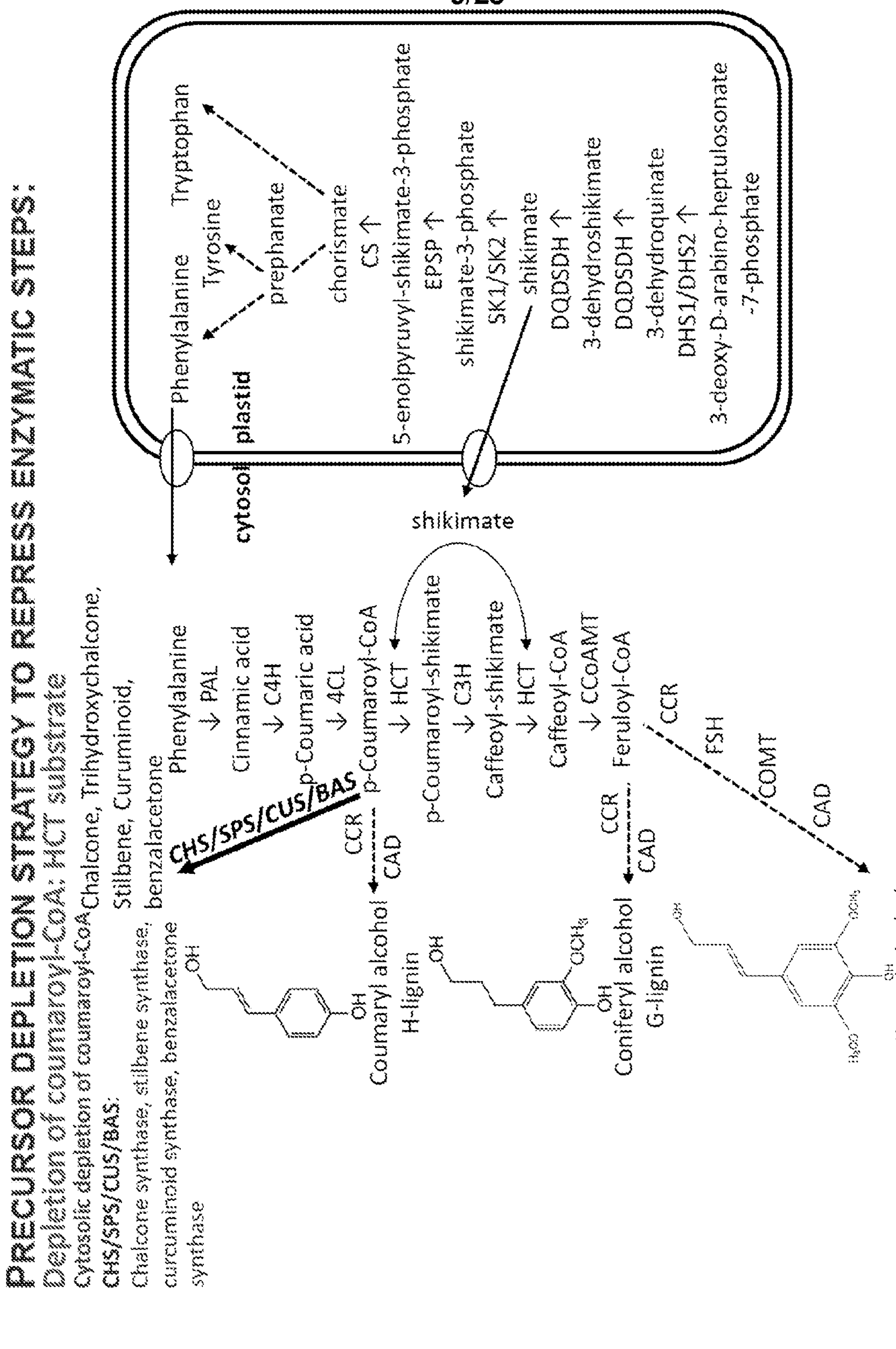

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of coumaroyl-CoA: HCT substrate
Cytosolic depletion of coumaroyl-CoA
RbC2'H: Ruta graveolens C2'H (2-oxoglutarate-dependent dioxygenase)
volatiles
Umbelliferone
RgC2'H
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
CAD
Coumaryl alcohol
H-lignin
Coniferyl alcohol
G-lignin
FSH
COMT
Sinapyl alcohol
S-lignin
shikimate
cytosol
plastid
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate
-7-phosphate

FIG. 8



PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of coumaroyl-CoA; HCT substrate
Cytosolic depletion of coumaroyl-CoA
CHS/SPS/CUS/BAS:
Chalcone synthase, stilbene synthase, curcuminoid synthase, benzalacetone synthase
Chalcone, Trihydroxychalcone, Stilbene, Curuminoid, benzalacetone
cytosol
plastid
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate -7-phosphate
shikimate
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CHS/SPS/CUS/BAS
CCR
CAD
CCR
CAD
CCR
F5H
COMT
CAD
OH
OH
Coumaryl alcohol
H-lignin
OH
OCH3
OH
Coniferyl alcohol
G-lignin
OH
OCH3
OH
H3CO
Sinapyl alcohol
S-lignin

FIG. 9

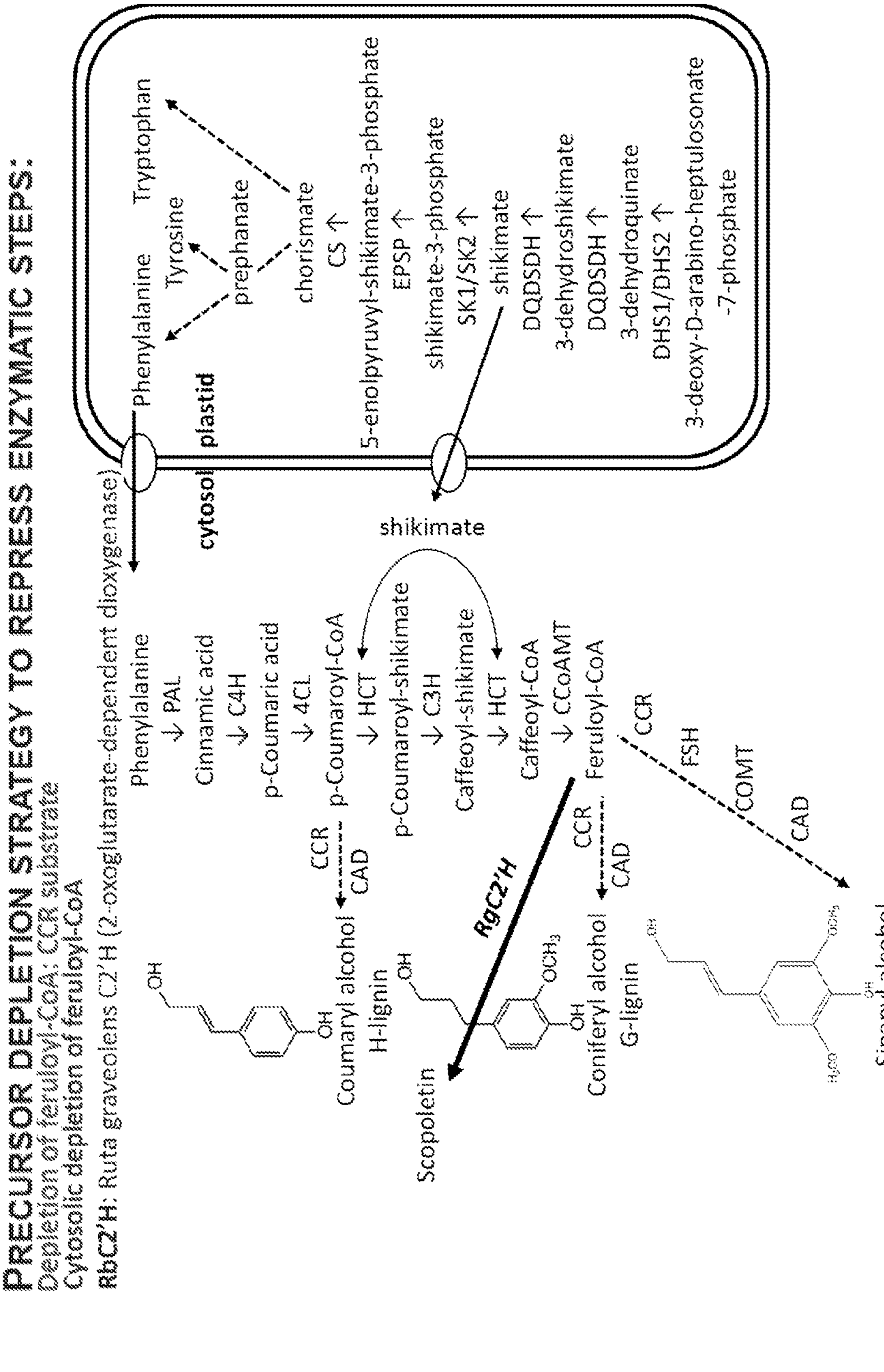

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of feruloyl-CoA: CCR substrate
Cytosolic depletion of feruloyl-CoA
RbC2'H: Ruta graveolens C2'H (2-oxoglutarate-dependent dioxygenase)
cytosol
plastid
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate
-7-phosphate
shikimate
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
FSH
COMT
CAD
CCR
CAD
CCR
CAD
RgC2'H
Coumaryl alcohol
H-lignin
Scopoletin
Coniferyl alcohol
G-lignin
Sinapyl alcohol
S-lignin

FIG. 10

11/28

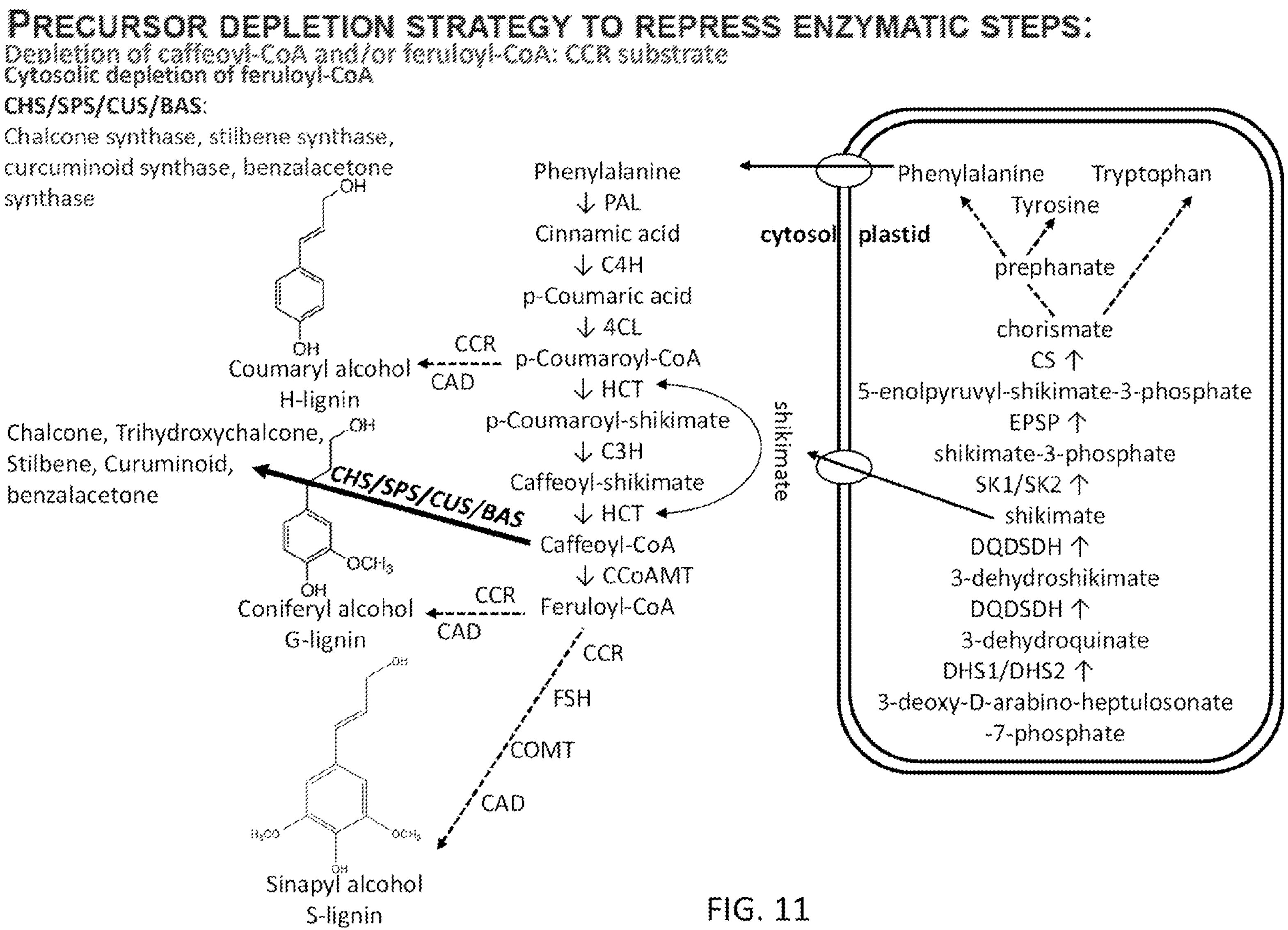

PRECURSOR DEPLETION STRATEGY TO REPRESS ENZYMATIC STEPS:
Depletion of caffeoyl-CoA and/or feruloyl-CoA: CCR substrate
Cytosolic depletion of feruloyl-CoA
CHS/SPS/CUS/BAS:
Chalcone synthase, stilbene synthase, curcuminoid synthase, benzalacetone synthase
Phenylalanine
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl-CoA
HCT
p-Coumaroyl-shikimate
C3H
Caffeoyl-shikimate
HCT
Caffeoyl-CoA
CCoAMT
Feruloyl-CoA
CCR
FSH
COMT
CAD
CCR
CAD
Coumaryl alcohol
H-lignin
Chalcone, Trihydroxychalcone, Stilbene, Curuminoid, benzalacetone
CHS/SPS/CUS/BAS
Coniferyl alcohol
G-lignin
Sinapyl alcohol
S-lignin
shikimate
cytosol
plastid
Phenylalanine
Tyrosine
Tryptophan
prephanate
chorismate
CS
5-enolpyruvyl-shikimate-3-phosphate
EPSP
shikimate-3-phosphate
SK1/SK2
shikimate
DQDSDH
3-dehydroshikimate
DQDSDH
3-dehydroquinate
DHS1/DHS2
3-deoxy-D-arabino-heptulosonate
-7-phosphate

FIG. 11

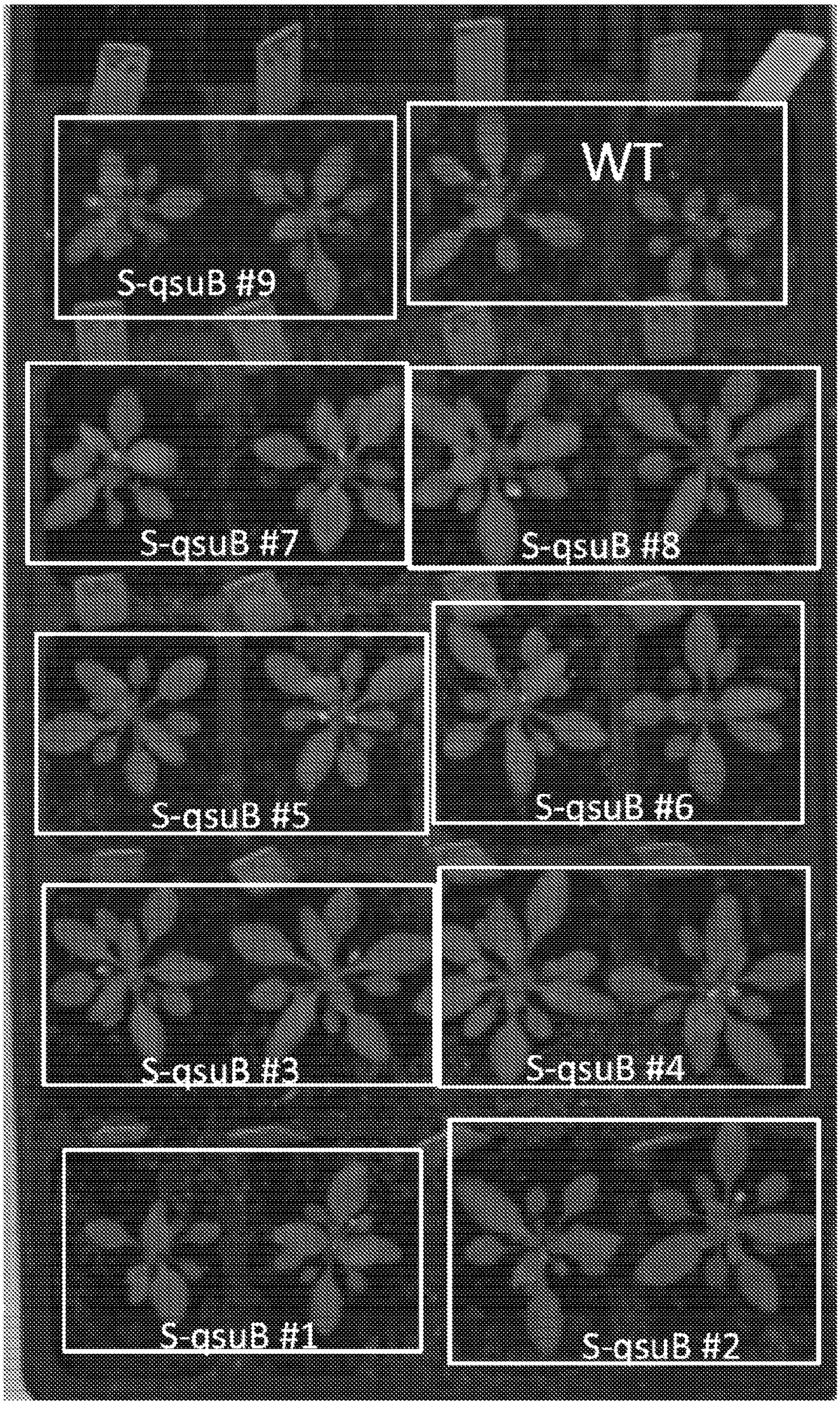

WT
S-qsuB #9
S-qsuB #7
S-qsuB #8
S-qsuB #5
S-qsuB #6
S-qsuB #3
S-qsuB #4
S-qsuB #1
S-qsuB #2

FIG. 12

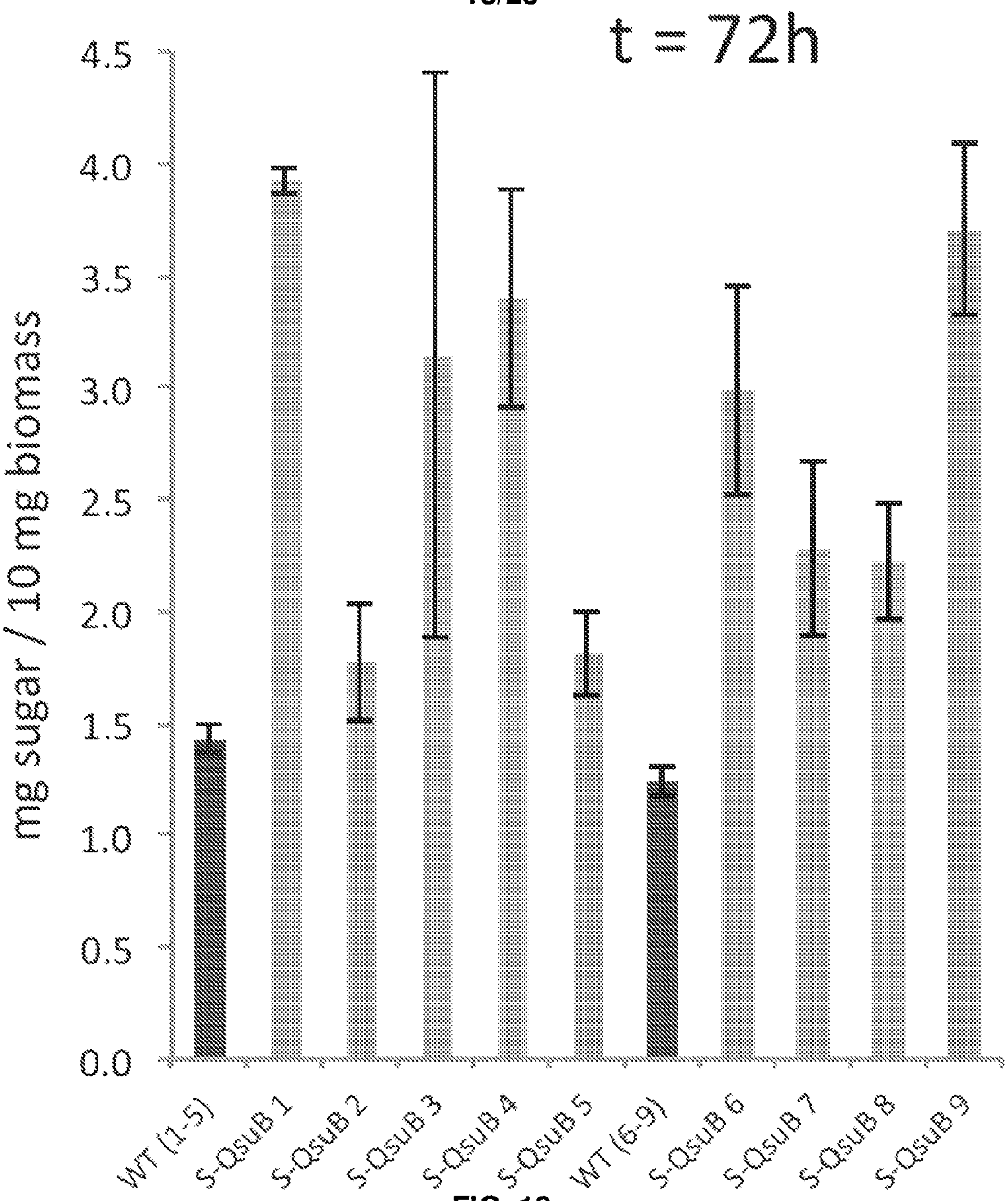

t = 72h
mg sugar / 10 mg biomass
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
WT (1-5)
S-QsuB 1
S-QsuB 2
S-QsuB 3
S-QsuB 4
S-QsuB 5
WT (6-9)
S-QsuB 6
S-QsuB 7
S-QsuB 8
S-QsuB 9

FIG. 13

FIG. 14

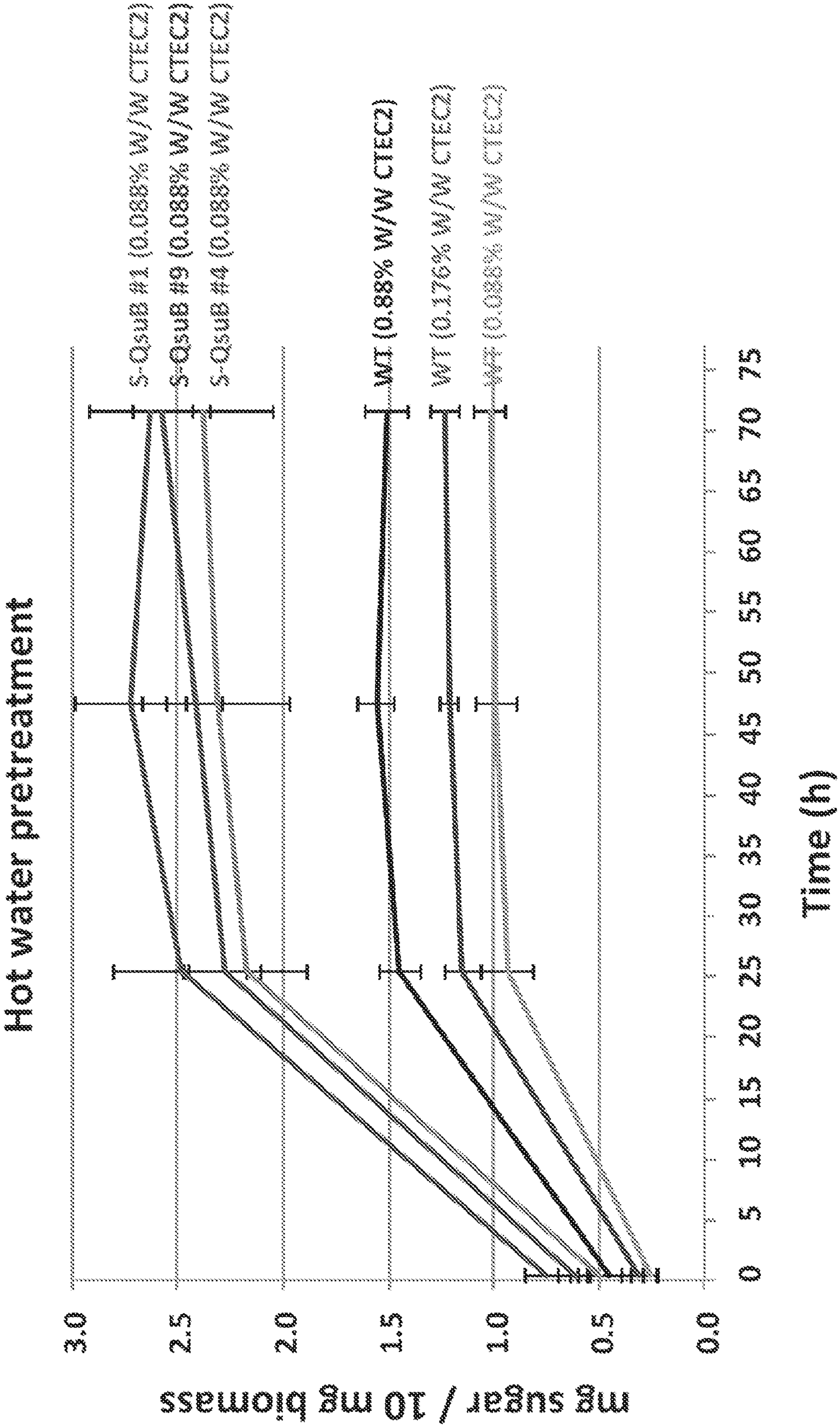

Hot water pretreatment
mg sugar / 10 mg biomass
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
Time (h)
S-QsuB #1 (0.088% W/W CTEC2)
S-QsuB #9 (0.088% W/W CTEC2)
S-QsuB #4 (0.088% W/W CTEC2)
WT (0.88% W/W CTEC2)
WT (0.176% W/W CTEC2)
WT (0.088% W/W CTEC2)

FIG. 15

Hot water pretreatment

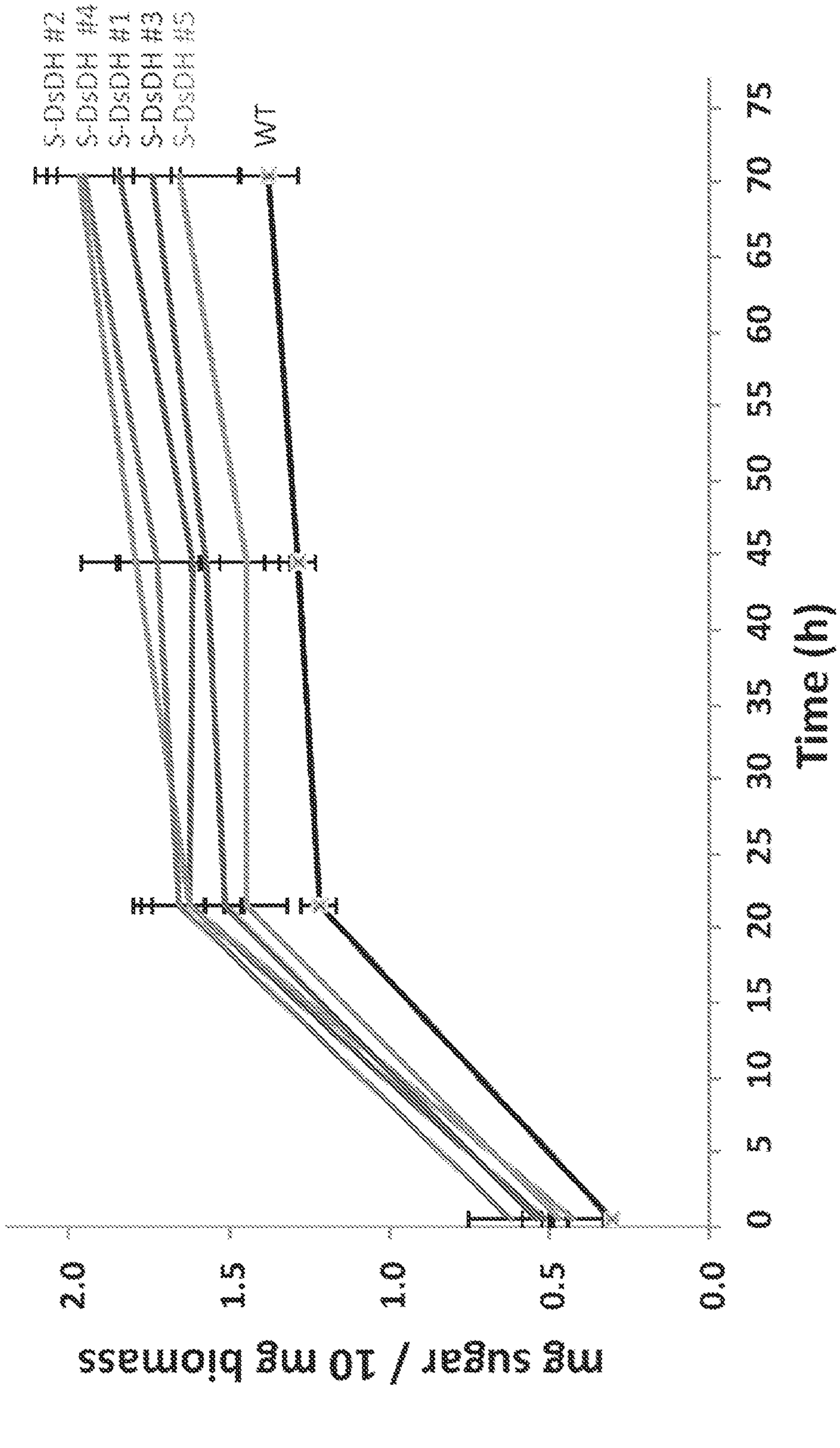
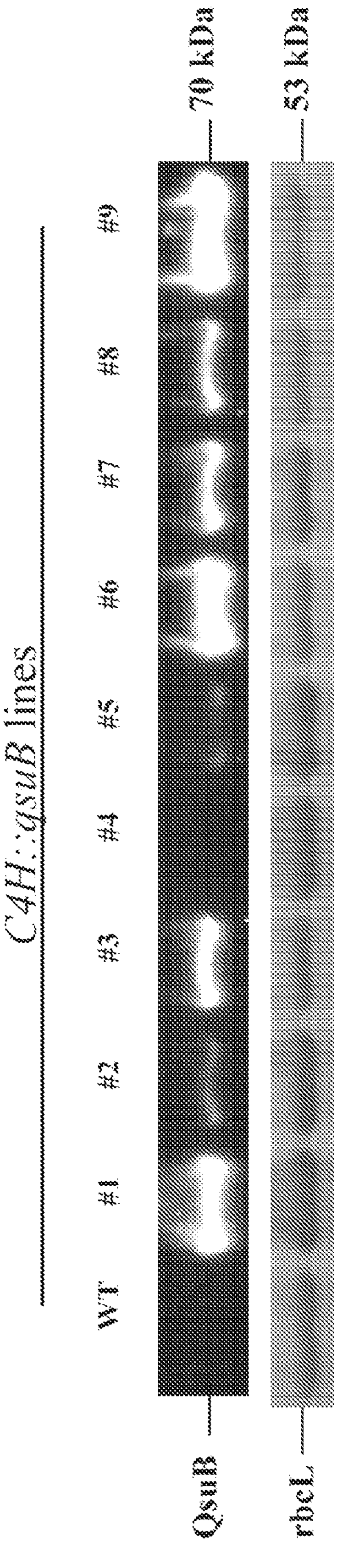


FIG. 16



17/28

FIG. 17

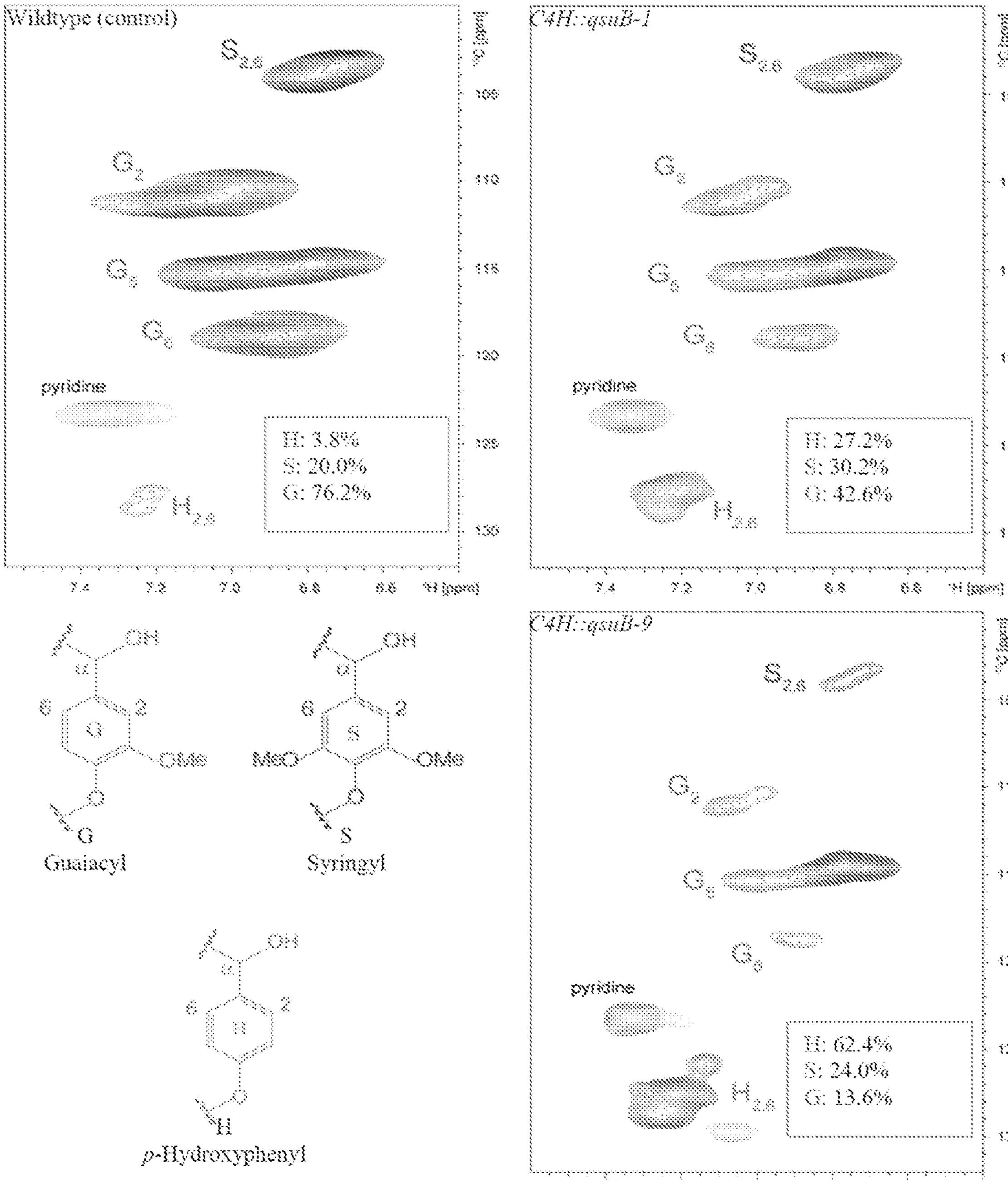

Wildtype (control)
S2,6
G2
G5
G6
pyridine
H2,6
H: 3.8%
S: 20.0%
G: 76.2%
C4H::qsuB-1
H: 27.2%
S: 30.2%
G: 42.6%
C4H::qsuB-9
H: 62.4%
S: 24.0%
G: 13.6%
Guaiacyl
Syringyl
p-Hydroxyphenyl
OMe
MeO
OH

FIG. 18

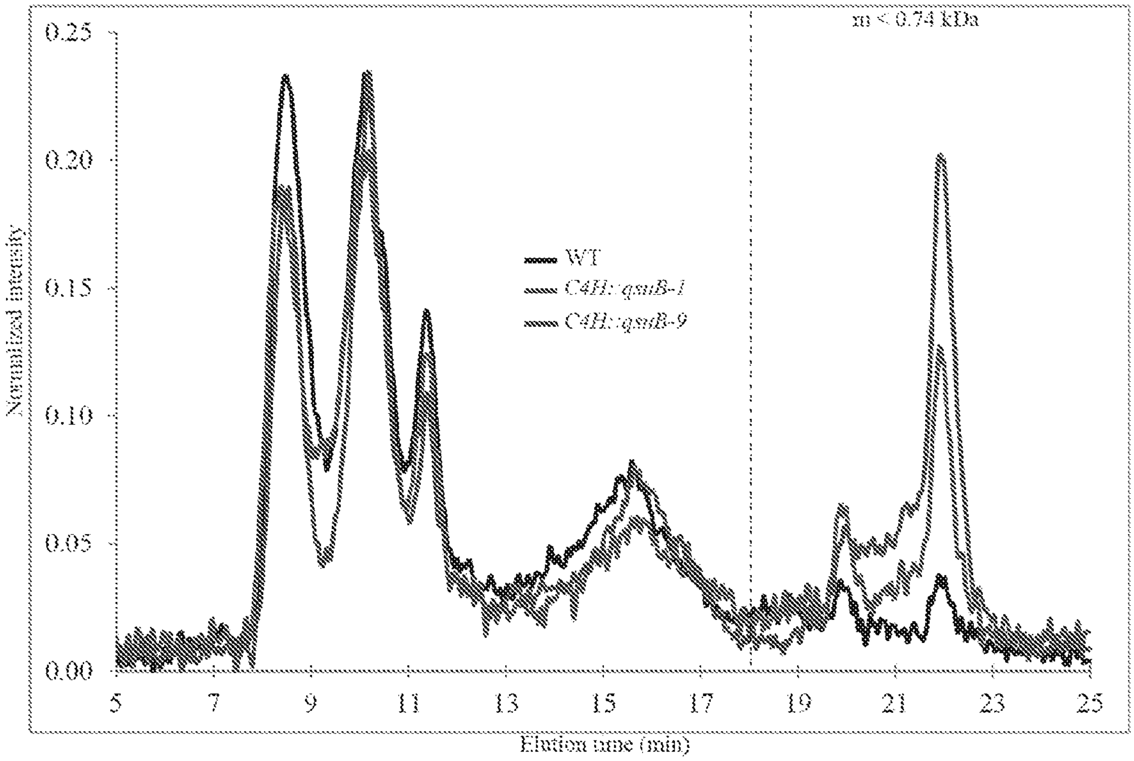
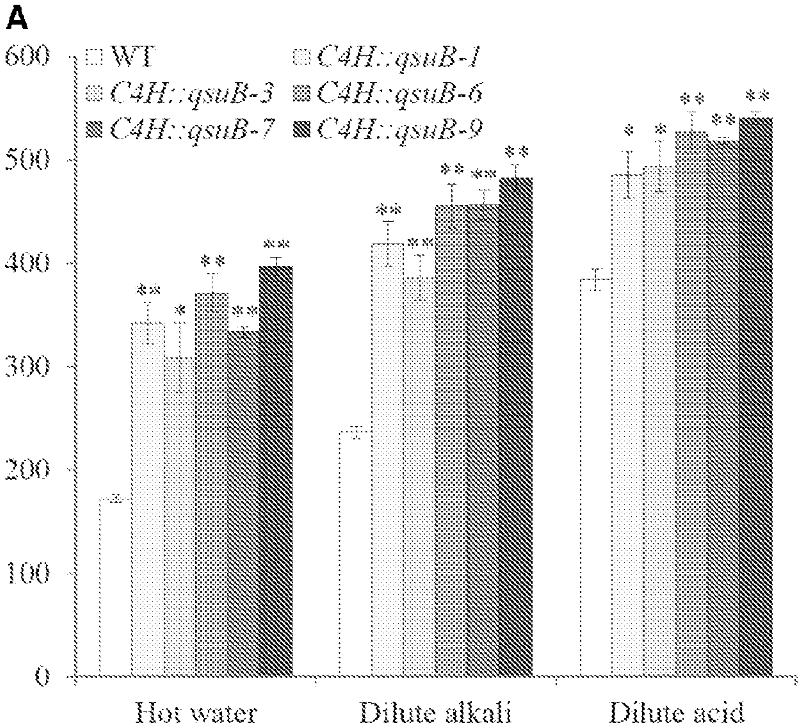
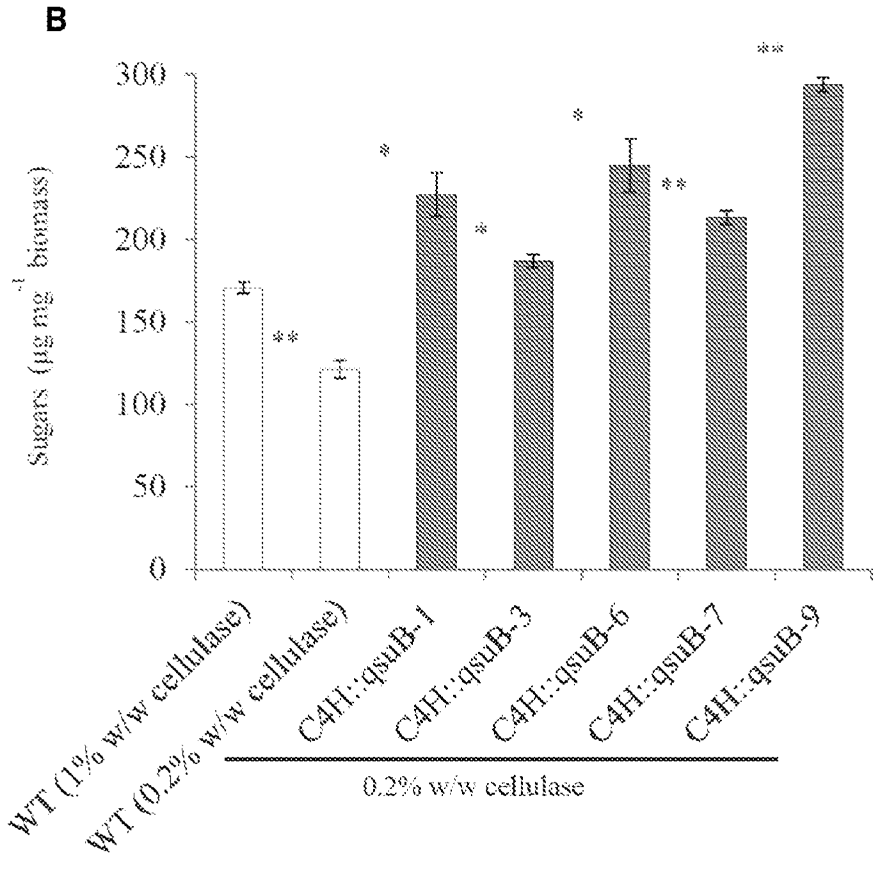

m < 0.74 kDa
WT
C4H::qsmB-1
C4H::qsmB-9
Normalized intensity
0.25
0.20
0.15
0.10
0.05
0.00
5
7
9
11
13
15
17
19
21
23
25
Elution time (min)

FIG. 19


A
WT
C4H::qsuB-1
C4H::qsuB-3
C4H::qsuB-6
C4H::qsuB-7
C4H::qsuB-9
600
500
400
300
200
100
0
Hot water
Dilute alkali
Dilute acid


B
Sugars (μg mg⁻¹ biomass)
300
250
200
150
100
50
0
WT (1% w/w cellulase)
WT (0.2% w/w cellulase)
C4H::qsuB-1
C4H::qsuB-3
C4H::qsuB-6
C4H::qsuB-7
C4H::qsuB-9
0.2% w/w cellulase

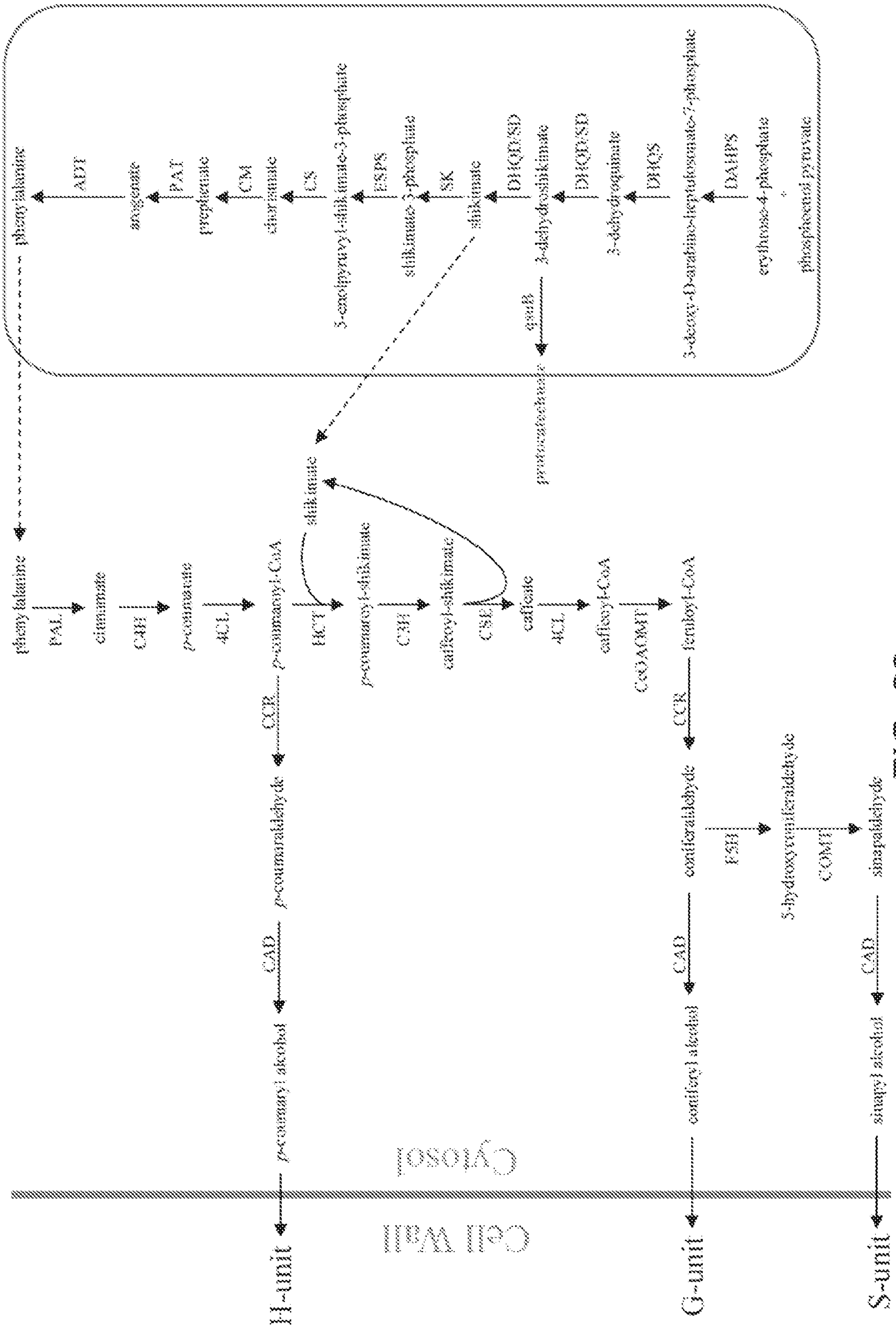

phenylalanine
ADT
arogenate
PAT
prephenate
CM
chorismate
CS
3-enolpyruvyl-shikimate-3-phosphate
ESPS
shikimate-3-phosphate
SK
shikimate
DHQD/SD
3-dehydroshikimate
DHQD/SD
3-dehydroquinate
DHQS
3-deoxy-D-arabino-heptulosonate-7-phosphate
DAHPS
erythrose-4-phosphate
+
phosphoenol pyruvate
quiB
protocatechuate
shikimate
phenylalanine
PAL
cinnamate
C4H
p-coumarate
4CL
p-coumaroyl-CoA
HCT
p-coumaroyl-shikimate
C3H
caffeoyl-shikimate
CSE
caffeate
4CL
caffeoyl-CoA
CcOAOMT
feruloyl-CoA
CCR
p-coumaraldehyde
CCR
coniferaldehyde
F5H
5-hydroxyconiferaldehyde
COMT
sinapaldehyde
CAD
p-coumaryl alcohol
CAD
coniferyl alcohol
CAD
sinapyl alcohol
Cytosol
Cell Wall
H-unit
G-unit
S-unit

FIG. 20

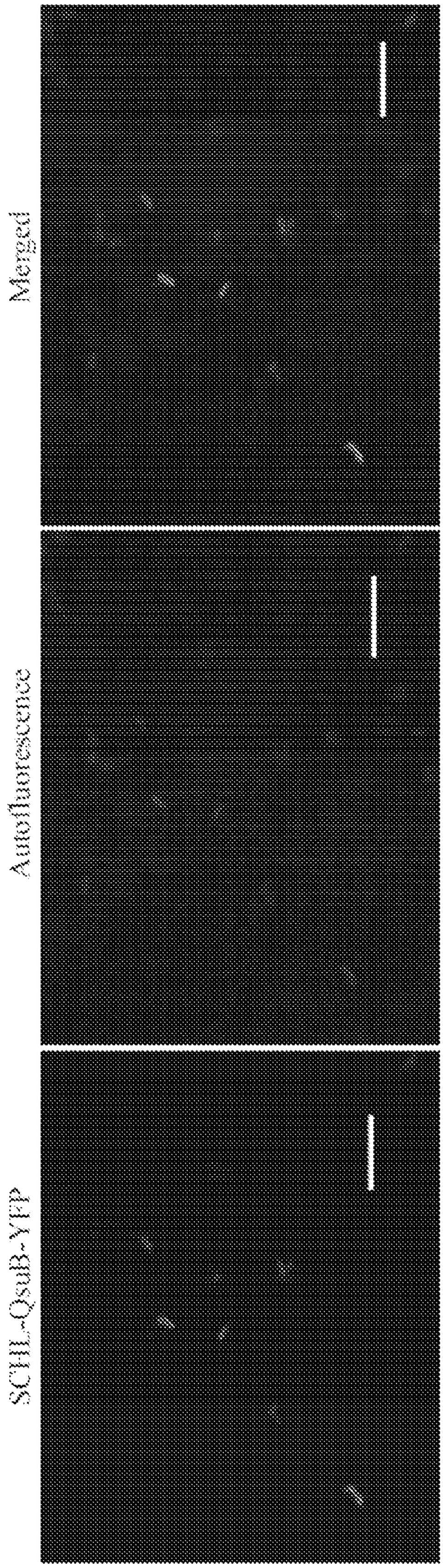


FIG. 21

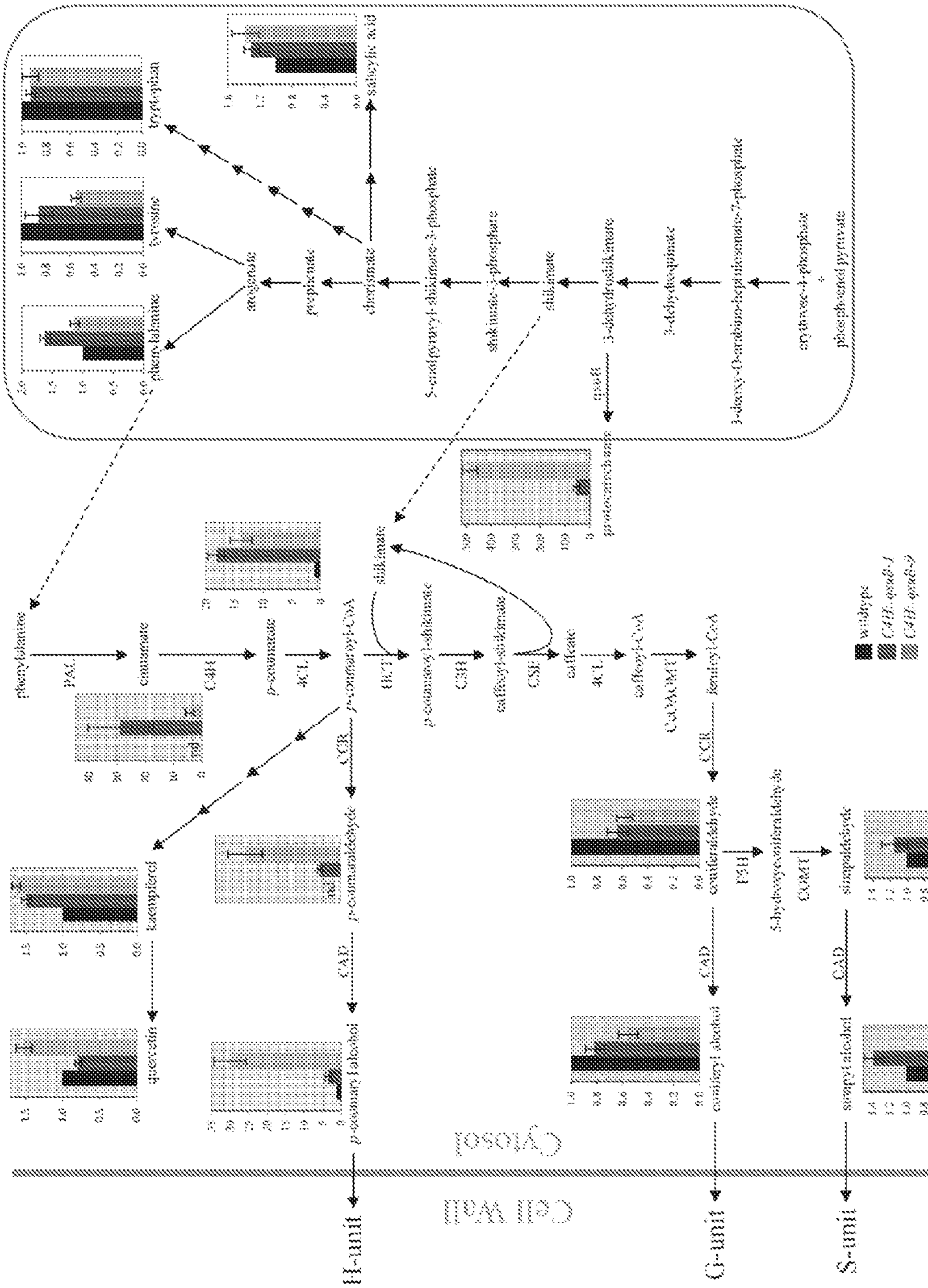


FIG. 22

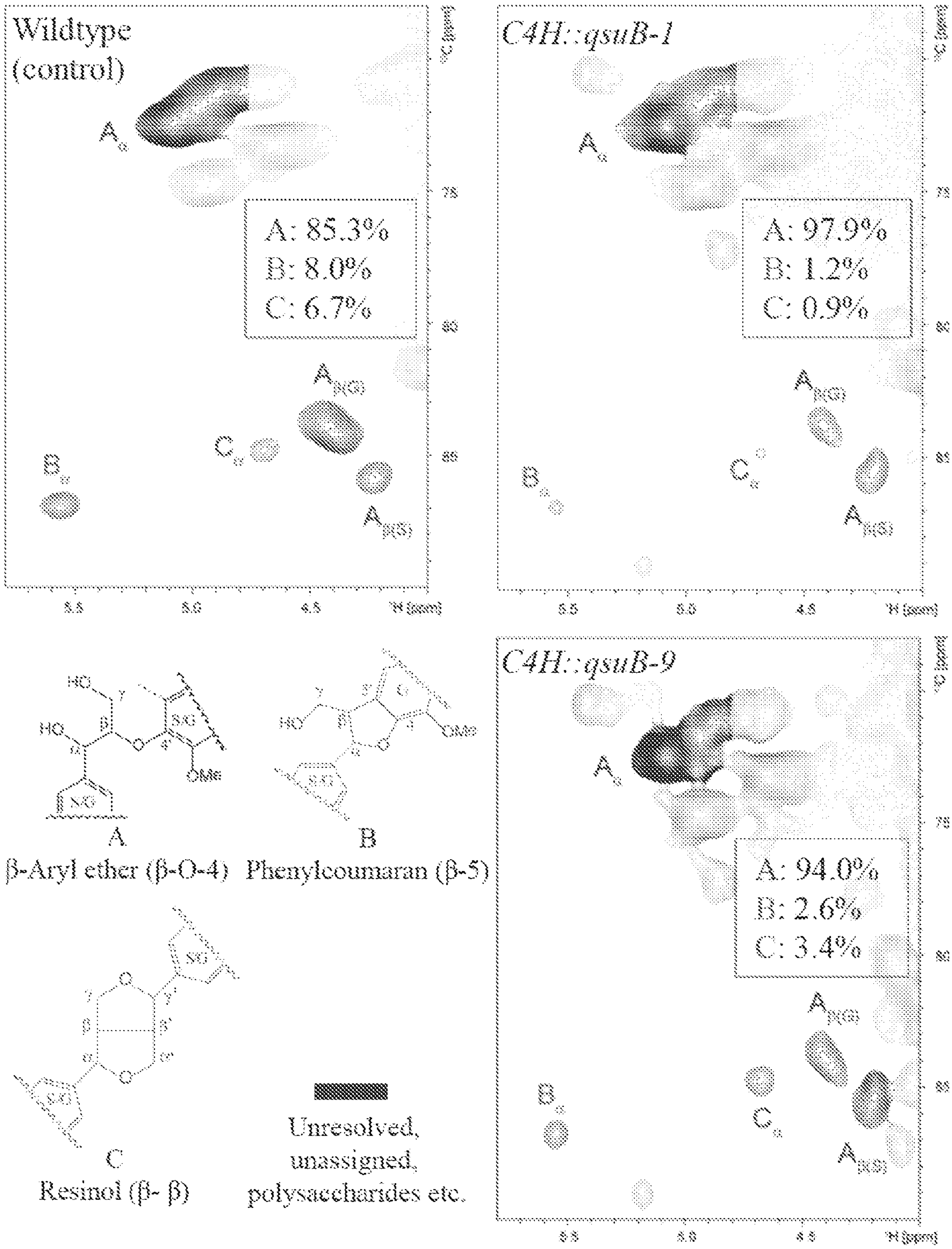

Wildtype
(control)
A: 85.3%
B: 8.0%
C: 6.7%
C4H::qsuB-1
A: 97.9%
B: 1.2%
C: 0.9%
C4H::qsuB-9
A: 94.0%
B: 2.6%
C: 3.4%
A
β-Aryl ether (β-O-4)
B
Phenylcoumaran (β-5)
C
Resinol (β- β)
Unresolved,
unassigned,
polysaccharides etc.

FIG. 23

24/28

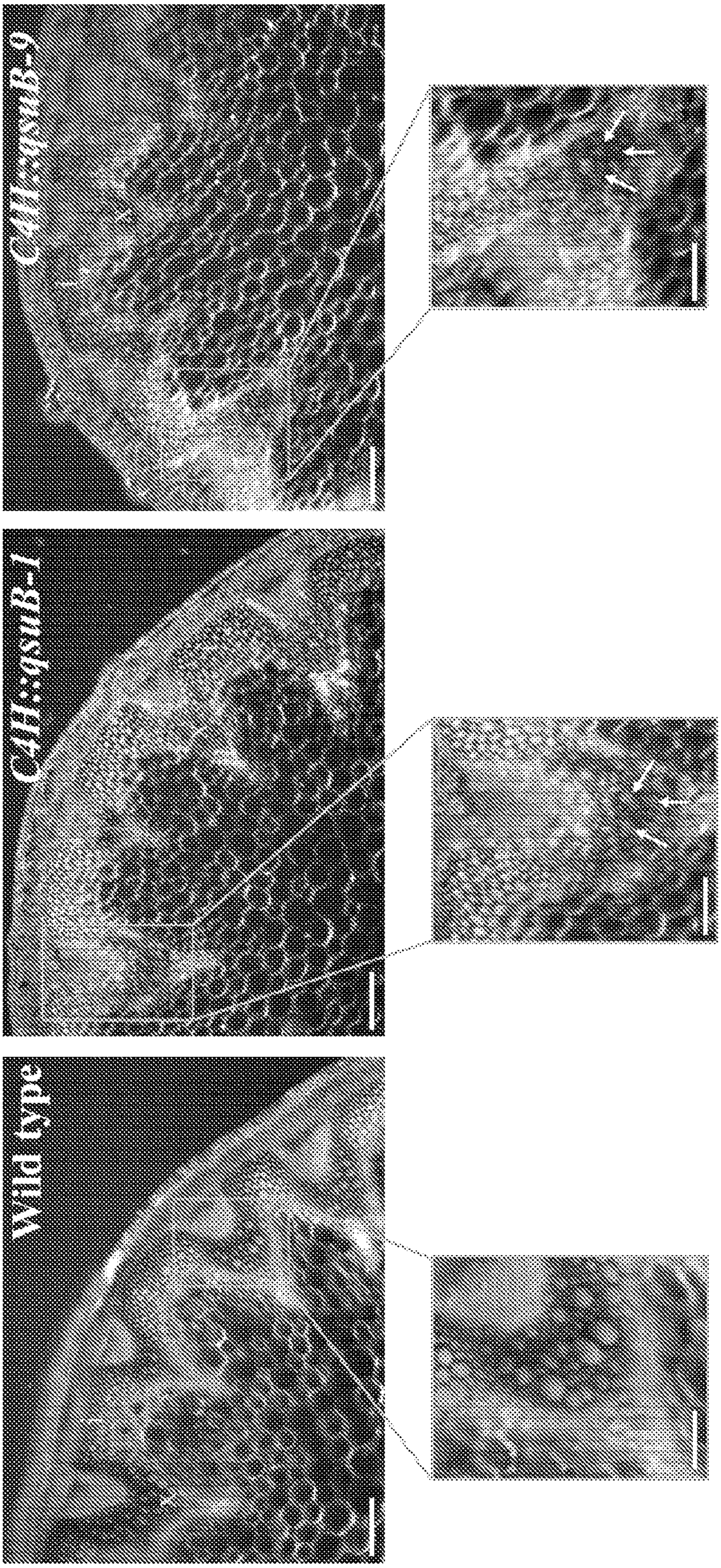


FIG. 24

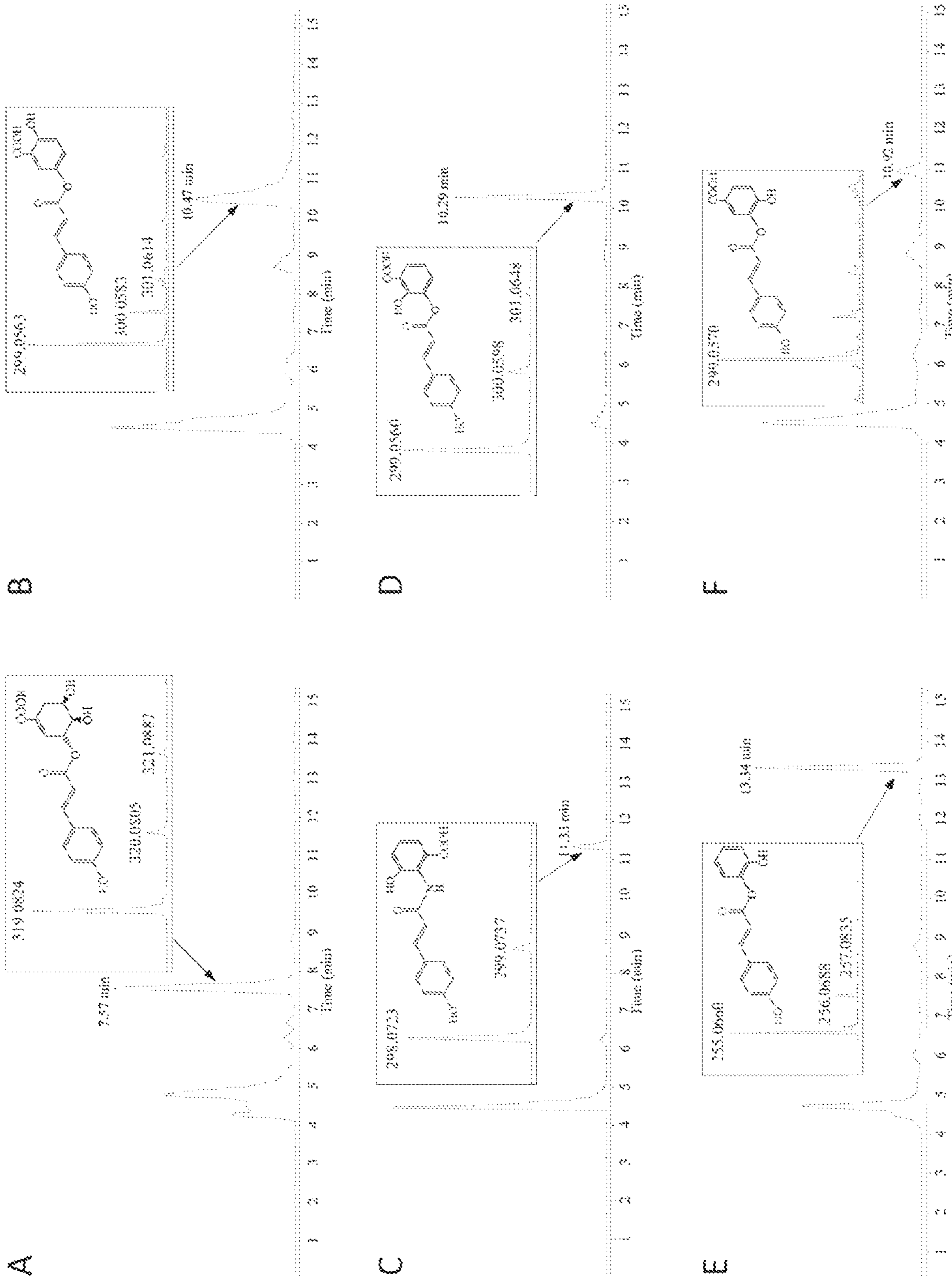


FIG. 25

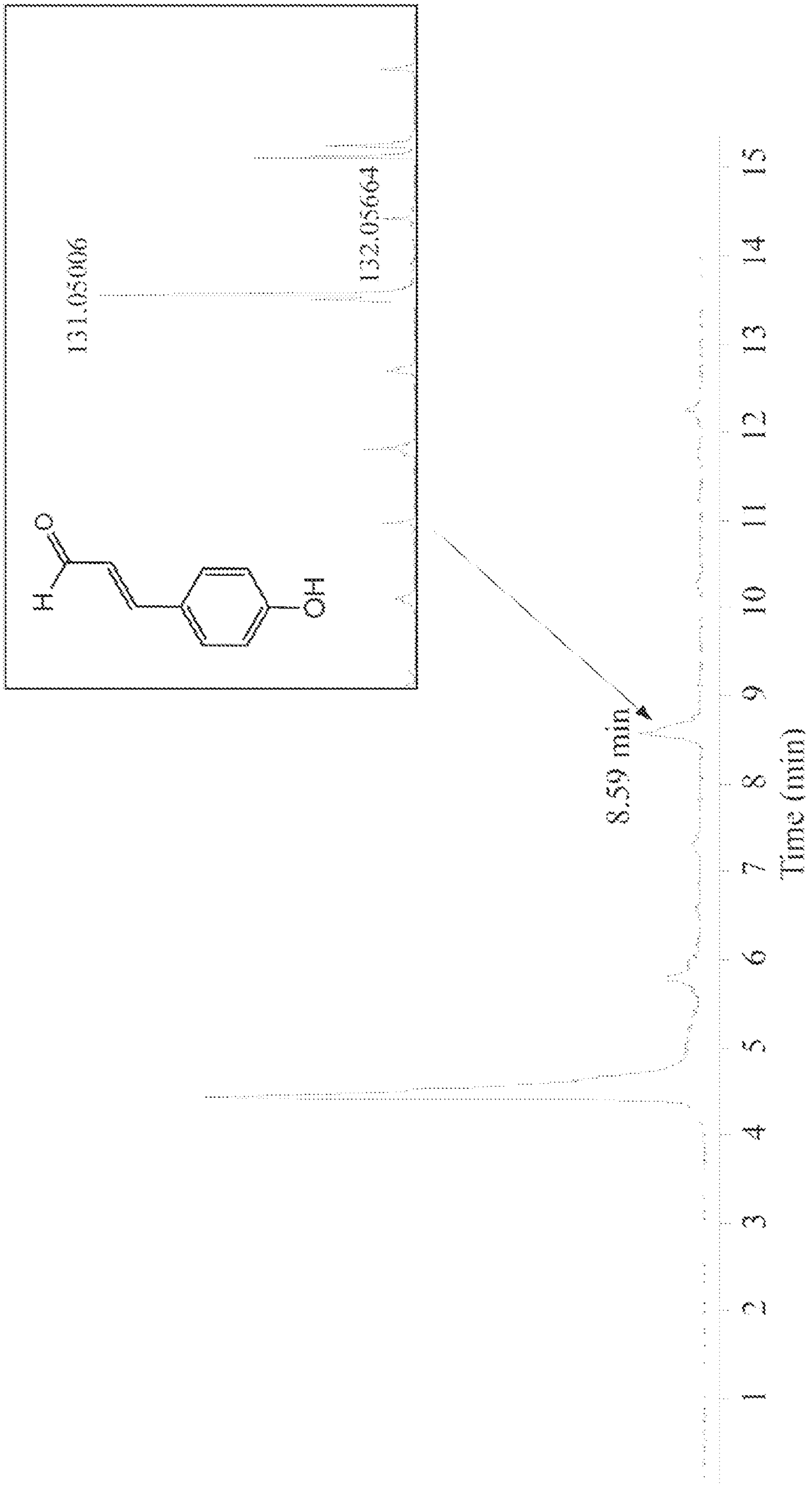

131.05006
132.05664
H
O
OH
8.59 min
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Time (min)

FIG. 26

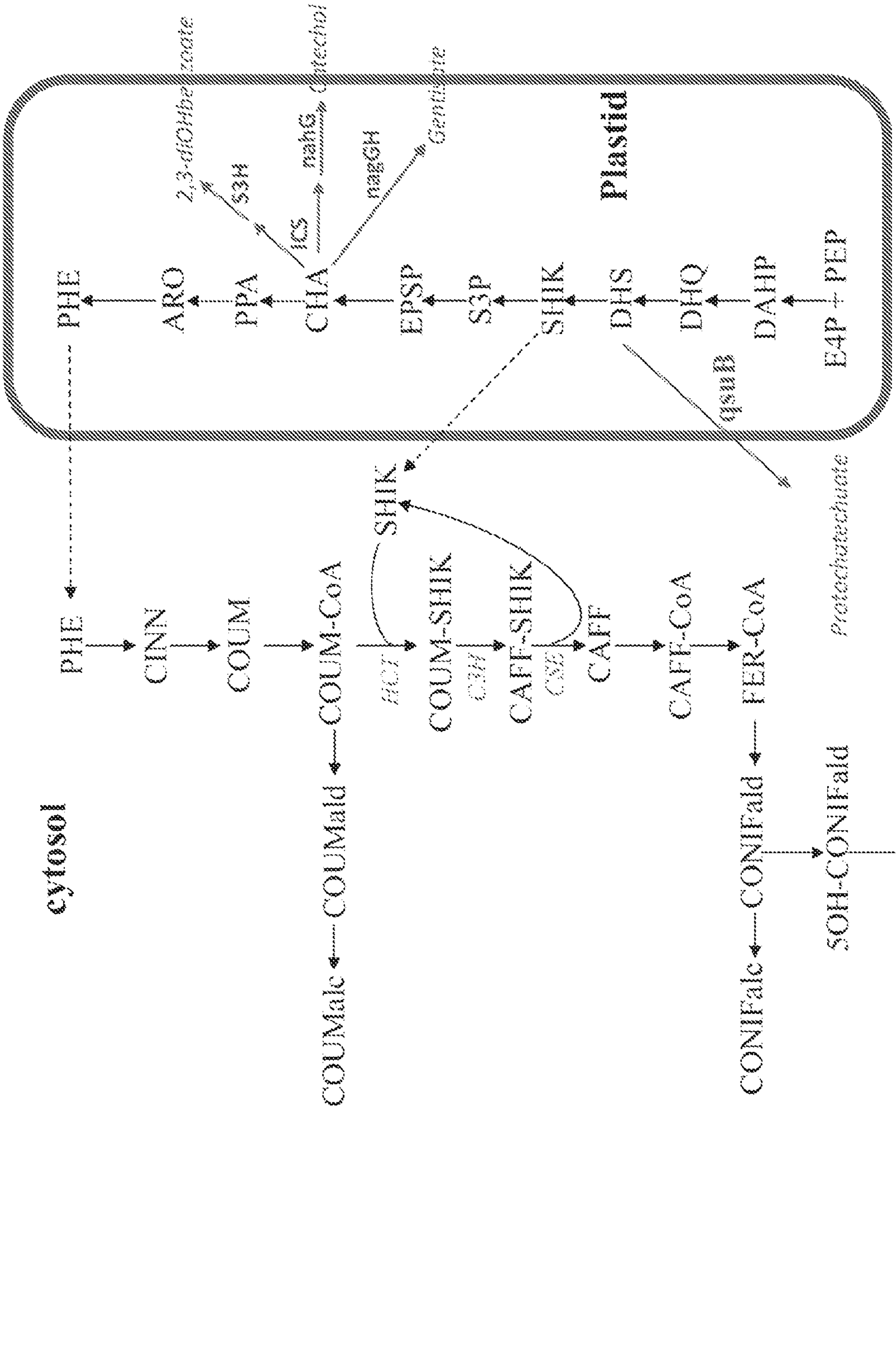
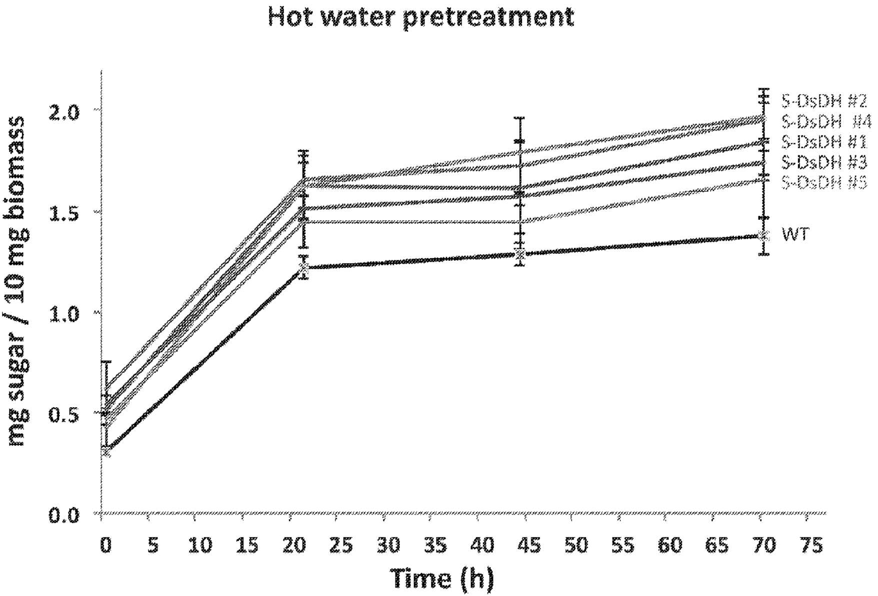

Plastid
cytosol
PHE
ARO
PPA
CHA
EPSP
S3P
SHIK
DHS
DHQ
DAHP
E4P + PEP
S3H
ICS
nahG
nagGH
2,3-diOHbenzoate
Catechol
Gentisate
qsuB
Protocatechuate
CINN
COUM
COUM-CoA
HCT
COUM-SHIK
C3H
CAFF-SHIK
CSE
CAFF
CAFF-CoA
FER-CoA
COUMald
COUMalc
CONIFald
CONIFalc
5OH-CONIFald
SINAPald
SINAPalc

FIG. 27

FIG. 28

Compound name	Origin	Formula	Molecular mass	Main mass fragments	Elution time (min)	WT (%)	*C4H::qsuB-1* (%)	*C4H::qsuB-9* (%)
Phenol	H	C_6H_6O	94	65, 66, *94*	4.28	1.3 (0.1)	5.1 (0.1)	10.5 (0.3)
2-Methylphenol	H	C_7H_8O	108	77, *107*, 108	4.96	0.6 (0.0)	1.5 (0.1)	5.7 (0.0)
3-Methylphenol	H	C_7H_8O	108	77, *107*, 108	5.16	1.4 (0.2)	4.6 (0.5)	10.9 (0.9)
2-Methoxyphenol	G	$C_7H_8O_2$	124	81, *109*, 124	5.34	5.6 (0.0)	3.7 (0.1)	3.1 (0.4)
2,5-Dimethylphenol	H	$C_8H_{10}O$	122	77, 107, *122*	5.93	0.4 (0.1)	1.7 (0.1)	4.1 (0.2)
4-Ethylphenol	H	$C_8H_{10}O$	122	77, *107*, 122	6.15	nd	nd	5.3 (0.3)
2-Methoxy-5-methylphenol	G	$C_8H_{10}O_2$	138	95, *123*, 138	6.45	7.6 (0.3)	9.0 (0.8)	4.0 (0.1)
4-Ethyl-2-methoxyphenol	G	$C_9H_{12}O_2$	152	122, *137*, 152	7.45	3.3 (0.1)	1.6 (0.2)	4.2 (0.1)
4-Ethenyl-2-methoxyphenol	G	$C_9H_{10}O_2$	150	107, *135*, 150	7.88	18.9 (0.4)	18.0 (0.2)	12.3 (0.2)
2,6-Dimethoxyphenol	S	$C_8H_{10}O_3$	154	111, 139, *154*	8.36	2.1 (0.1)	1.4 (0.1)	4.4 (0.3)
2-Methoxy-4-propenylphenol	G	$C_{10}H_{12}O_2$	164	131, 149, *164*	8.41	3.3 (0.0)	2.6 (0.6)	nd
4-hydroxy-3-methoxyphenylacetaldehyde	G	$C_{10}H_{14}O_2$	166	122, *137*, 166	8.52	0.4 (0.0)	0.6 (0.0)	nd
4-Hydroxy-3-methoxybenzaldehyde	G	C_8H8O3	152	109, 151, *152*	9.02	8.6 (0.3)	nd	0.3 (0.1)
4-Methyl-2,6-dimethoxyphenol	S	$C_9H_{12}O_3$	168	125, 153, *168*	9.47	3.1 (0.1)	3.1 (0.0)	1.9 (0.1)
2-Methoxy-4-propenylphenol	G	$C_{10}H_{12}O_2$	164	131, 149, *164*	9.52	11.4 (0.2)	8.9 (0.2)	5.3 (0.4)
4-Ethyl-2,6-dimethoxyphenol	S	$C_{10}H_{14}O_3$	182	*167*, 182	10.42	1.1 (0.1)	0.9 (0.3)	2.2 (0.1)
4-Hydroxy-3-methoxyphenyl acetone	G	$C_{10}H_{12}O_3$	180	122, *137*, 180	10.56	2.1 (0.1)	nd	nd
4-Hydroxy-3,5-dimethoxystyrene	S	$C_{10}H_{12}O_3$	180	137, 165, *180*	10.88	10.5 (0.3)	16.5 (0.5)	12.7 (0.2)
4-Allyl-2,6-dimetoxyphenol	S	$C_{11}H_{14}O_3$	194	167, 179, *194*	11.32	2.3 (0.2)	3.6 (0.3)	1.0 (0.3)
4-Hydroxy-3,5-dimethoxybenzaldehyde	S	$C_9H_{10}O_4$	182	167, 181, *182*	12.07	2.6 (0.1)	nd	1.4 (0.3)
4-Propinyl-2,6-dimethoxyphenol	S	$C_{11}H_{12}O_3$	192	106, 131, 177, *192*	12.23	0.9 (0.1)	1.1 (0.4)	nd
4-Propenyl-2,6-dimethoxyphenol	S	$C_{11}H_{14}O_3$	194	167, 179, *194*	12.43	7.8 (0.5)	9.1 (0.2)	6.6 (1.2)
4-Hydroxy-3,5-dimethoxyacetophenone	S	$C_{10}H_{12}O_4$	196	153, *181*, 196	12.88	0.8 (0.1)	0.9 (0.6)	0.4 (0.2)
4-hydroxy-3-methoxycinnamaldehyde	G	$C_{10}H_{10}O_3$	178	107, 135, 147, *178*	13.00	2.2 (0.4)	nd	nd
4-hydroxy-3,5-dimethoxyphenylacetone	S	$C_{11}H_{14}O_4$	210	123, *167*, 210	13.26	1.3 (0.0)	2.2 (0.2)	2.9 (0.5)
4-Hydroxy-3,5-dimethoxyphenylethanone	S	$C_{10}H_{12}O_4$	196	153, *181*, 196	13.83	0.5 (0.0)	nd	nd
% H-units						3.7 (0.2)	13.2 (0.7)	36.8 (1.3)
% G-units						63.3 (0.3)	46.8 (0.9)	29.3 (0.3)
% S-units						32.7 (0.7)	40.0 (0.2)	33.9 (1.0)


Hot water pretreatment
mg sugar / 10 mg biomass
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
Time (h)
S-DsDH #2
S-DsDH #4
S-DsDH #1
S-DsDH #3
S-DsDH #5
WT