**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE LORENA**

GUILHERME RODRIGUES FERRAZ

Identificação, anotação e análise filogenética das famílias gênicas envolvidas na via de biossíntese de lignina em cana-de-açúcar (*Saccharum spp.*)

**2016
Lorena – SP**

GUILHERME RODRIGUES FERRAZ

Identificação, anotação e análise filogenética das famílias gênicas envolvidas na via de biossíntese de lignina em cana-de-açúcar (*Saccharum spp.*)

Dissertação apresentada à Escola de Engenharia de Lorena da Universidade de São Paulo para obtenção do título de Mestre em Ciências do Programa de Pós-Graduação em Biotecnologia Industrial na área de conversão de biomassa.

Orientador: Prof. Dr. Elisson Antonio da Costa Romanel

Versão Original

2016
Lorena – SP

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE

Ficha catalográfica elaborada pelo Sistema Automatizado
da Escola de Engenharia de Lorena,
com os dados fornecidos pelo(a) autor(a)

Ferraz, Guilherme Rodrigues

Identificação, anotação e análise filogenética das famílias gênicas envolvidas na via de biossíntese de lignina em cana-de-açúcar (*Saccharum spp.*) / Guilherme Rodrigues Ferraz; orientador Elisson Antonio da Costa Romanel - Versão Original. - Lorena, 2016.
89p.

Dissertação (Mestrado em Ciências - Programa de Pós Graduação em Biotecnologia Industrial na Área de Conversão de Biomassa) - Escola de Engenharia de Lorena da Universidade de São Paulo. 2016
Orientador: Elisson Antonio da Costa Romanel

1. Cana-de-açúcar. 2. Lignina. 3. Bioinformática.
I. Título. II. Romanel, Elisson Antônio da Costa, orient.

RESUMO

FERRAZ, G. R. **Identificação, anotação e análise filogenética das famílias gênicas envolvidas na via de biossíntese de lignina em cana-de-açúcar (*Saccharum spp.*)**. 2016. 89p. Dissertação (Mestrado em Ciências) – Universidade de São Paulo, Escola de Engenharia de Lorena, Lorena, 2016.

A parede celular vegetal é composta primariamente de celulose, hemicelulose e lignina. A lignina é o segundo biopolímero mais abundante na biosfera, atrás apenas de celulose, e é formado principalmente a partir da ligação entre três monômeros chamados monolignóis. A formação desses monolignóis é catalisada por pelo menos 10 enzimas membros das famílias gênicas AMP_binding (gene *4CL*), p450 (*C3H*, *C4H* e *F5H*), ADH_N (*CAD*), Epimerase (*CCR*), Methyltransf_3 (*CCoAOMT*), Methyltransf_2 (*COMT*), Transferase (*HCT*) e Lyase_aromatic (*PAL*), que compõem a via de biossíntese dos monolignóis. Até o momento, cerca de 25 sequências da via de biossíntese de lignina já foram identificados em cana-de-açúcar (*Saccharum spp.*). Ainda, o sequenciamento do genoma desta espécie se mostra uma difícil tarefa devido ao tamanho do genoma e ploidia nuclear. Em virtude da disponibilidade de transcriptomas oriundos de diferentes órgãos de cana-de-açúcar e a importância na identificação correta das sequências ortólogas, o presente trabalho visa à identificação, anotação e análise filogenética dos genes das famílias gênicas envolvidas na biossíntese de lignina em cana-de-açúcar. Para isso, genes destas famílias gênicas da planta modelo de eudicotiledôneas *Arabidopsis thaliana* e gramíneas tais como *Oryza sativa* (arroz), *Brachypodium distachyon*, *Zea mays* (milho) e *Sorghum bicolor* (sorgo) foram identificados, anotados e filogeneticamente categorizados. Em seguida, as sequências de cana-de-açúcar foram identificadas em cinco transcriptomas distintos, anotados e avaliados quanto à identidade e cobertura em relação às proteínas de sorgo. As análises filogenômicas entre cana-de-açúcar e as demais espécies estudadas revelaram 23 sequências candidatas envolvidos na biossíntese de lignina em cana-de-açúcar.

Palavras-chave: Lignina. Cana-de-açúcar. Filogenia. Bioinformática.

ABSTRACT

FERRAZ, G. R. **Identification, annotation and phylogenetic analysis of gene families involved in lignin biosynthesis pathway in sugarcane (*Saccharum spp.*)**. 2016. 89p. Dissertation (Master of Science) – Escola de Engenharia de Lorena, Universidade de São Paulo, Lorena, 2016.

Plant cell wall is composed of cellulose, hemicellulose, and lignin. Lignin is the second most abundant biopolymer on biosphere, after only cellulose, and is formed primarily from three monomers, called monolignols. Monolignol biosynthesis is catalysed by at least 10 enzymes members of AMP_binding (4CL enzyme), p450 (C3H, C4H e F5H), ADH_N (CAD), Epimerase (CCR), Methyltransf_3 (CCoAOMT), Methyltransf_2 (COMT), Transferase (HCT) e Lyase_aromatic (PAL) gene families involved in the biosynthetic pathway of monolignols. So far, there is 25 sugarcane (*Saccharum spp.*) genes identified involved in the lignin biosynthesis. Sequencing sugarcane's genome is still a difficult task due to large genome and ploidy level. Due to the availability of transcriptome data from distinct organs of sugarcane and the importance in the accuracy on the identification of the orthologous sequences, this work aim the identification, annotation and phylogenetic analysis of proteins members of distinct gene family involved in monolignol biosynthesis in sugarcane. To conduct it, we identified and annotated all genes from the model plant eudicots *Arabidopsis thaliana* and grasses such as *Oryza sativa* (rice), *Brachypodium distachyon*, *Zea mays* (mayze) and *Sorghum bicolor* (sorghum) and analysed the phylogenic relationship among them. The sugarcane sequences were retrieved and annotated from five distinct transcriptome data and analysed by the identity and coverage against the best sorghum orthologous genes. The phylogenomics analysis revealed a total of 23 sugarcane sequences putatively involved in the biosynthesis of lignin.

Keywords: Lignin. Cell Wall. Sugarcane. Bioinformatics.

Lista de Figuras

Figura 1 - Diagrama esquemático da história evolutiva de gramíneas.....	8
Figura 2 - Interação entre os componentes da parede celular: polissacarídeos e lignina.....	10
Figura 3 - Diagrama da organização da parede celular vegetal.....	11
Figura 4 - Principais monolignóis.....	14
Figura 5 - Via de biossíntese de monolignóis.....	16
Figura 6 - Árvore filogenética das proteínas membro da família gênica AMP_binding e do clado 4CL nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	36
Figura 7 - Árvore filogenética das proteínas membro da família gênica p450 e dos clados C3H, C4H e F5H nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	37
Figura 8 - Árvore filogenética das proteínas membro da família gênica ADH_N e do clado CAD nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	38
Figura 9 - Árvore filogenética das proteínas membro da família gênica Methyltransf_3 e do clado CCoAOMT nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	39
Figura 10 - Árvore filogenética das proteínas membro da família gênica Epimerase (clados B e C) e do clado CCR nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i> ...	40
Figura 11 - Árvore filogenética das proteínas membro da família gênica Methyltransf_2 e do clado COMT nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	41
Figura 12 - Árvore filogenética das proteínas membro da família gênica Transferase e do clado HCT nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	42
Figura 13 - Árvore filogenética das proteínas membro da família gênica Lyase_aromatic, enzimas PAL, nas espécies <i>A. thaliana</i> , <i>B. distachyon</i> , <i>O. sativa</i> , <i>S. bicolor</i> e <i>Z. mays</i>	43
Figura 14 - Alinhamento de sequências proteicas de 4CL dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	47
Figura 15 - Alinhamento de sequências proteicas de C3H dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	52
Figura 16 - Alinhamento de sequências proteicas de C4H dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	54
Figura 17 - Alinhamento de sequências proteicas de CAD dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	57
Figura 18 - Alinhamento de sequências proteicas de CCoAOMT dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	57
Figura 19 - Alinhamento de sequências proteicas de CCR dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	58
Figura 20 - Alinhamento de sequências proteicas de COMT dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	60
Figura 21 - Alinhamento de sequências proteicas de F5H dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	61
Figura 22 - Alinhamento de sequências proteicas de HCT dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	62
Figura 23 - Alinhamento de sequências proteicas de PAL dos diferentes bancos de cana-de-açúcar com os genes <i>bona fide</i> de <i>S. bicolor</i>	63
Figura 24 - Árvores filogenéticas dos clados <i>bona fide</i> das enzimas envolvidas na biossíntese de lignina entre cana-de-açúcar, <i>Brachypodium</i> , arroz, milho e sorgo e <i>Arabidopsis</i>	65

Lista de Tabelas

Tabela 1 - Composição da parede celular vegetal em diferentes espécies.....	12
Tabela 2 - Espécies de interesse energético já estudadas.....	21
Tabela 3 - Número de genes das famílias gênicas relacionadas com a biossíntese de lignina.....	33
Tabela 4 - Número de genes das famílias gênicas relacionadas com a biossíntese de lignina nas diferentes bibliotecas de cana-de-açúcar.....	34
Tabela 5 - Número de genes candidatos em cana-de-açúcar e suas respectivas relações de identidade e cobertura com sequência ortólogo de sorgo.....	66

SUMÁRIO

1 INTRODUÇÃO	7
1.1 Evolução e genômica de plantas	7
1.2 Parede vegetal	9
1.3 Lignina	13
1.4 Genética da biossíntese de lignina	16
1.5 Cana-de-açúcar: uso como matéria-prima, genética e variedades	22
1.6 Bioinformática	25
1.7 Justificativa	27
2 OBJETIVOS	28
2.1 Objetivos gerais	28
2.2 Objetivos específicos	28
3 MATERIAL E MÉTODOS	29
3.1 Identificação e anotação das sequências pertencentes às famílias gênicas relacionadas à biossíntese de lignina em <i>Arabidopsis thaliana</i> , gramíneas e cana-de- açúcar.....	29
3.2 Alinhamentos e análises filogenéticas.....	30
4 RESULTADOS E DISCUSSÃO	32
4.1 Identificação e anotação dos genes de <i>Arabidopsis</i> e gramíneas.....	32
4.2 Identificação e anotação das sequências de cana-de-açúcar.....	34
4.3 Identificação dos clados bona fide da via de biossíntese de lignina em gramíneas...	35
4.4 Identificação das sequências de cana-de-açúcar pertencentes aos clados bona fide e definição dos genes candidatos.....	45
5 CONCLUSÃO	67
REFERÊNCIAS	68
ANEXO.....	89

1 INTRODUÇÃO

1.1 Evolução e Genômica de Plantas

As angiospermas representam um dos grupos mais diversos de plantas e variam em morfologia, tamanho e habitat (ILIC *et al.*, 2007). Com cerca de 250.000 espécies, estão presentes na maioria dos ecossistemas e representam a maior parte das plantas terrestres (WIKSTRÖM *et al.*, 2001). Se originaram no Médio-baixo Jurássico, cerca de 179-153 milhões de anos (WIKSTRÖM *et al.*, 2001) e há cerca de 130 milhões de anos surgiu o primeiro grande grupo de angiospermas, as monocotiledôneas (BREMER, 2002; CHASE *et al.* 2006).

A família das *Poaceae*, as gramíneas, inclui mais de 11000 espécies conhecidas com uma distribuição mundial (CLAYTON; RENVOIZE, 1986; GRASS PHYLOGENY WORKING GROUP II, 2012). Fazem parte dessa família algumas das mais importantes culturas agronômicas, como trigo (*Triticum aestivum*), arroz (*Oryza sativa*) e milho (*Zea mays*) (GRASS PHYLOGENY WORKING GROUP II, 2012), além da cana-de-açúcar e outras espécies com potencial para biocombustíveis (BYRT *et al.*, 2011), formando a chamada superfamília *Panicoidae* (Figura 1) (MATHEWS *et al.*, 2000).

A agricultura é quase completamente dependente das angiospermas (ILIC *et al.*, 2007), em especial da família *Poaceae*. Aproximadamente um terço da superfície terrestre está coberta com alguma espécie de *Poaceae*, e a maior parte da população humana depende fortemente de cereais de gramíneas para seu sustento diário (MATHEWS *et al.*, 2000).

Figura 1- Diagrama esquemático da história evolutiva de gramíneas.

Números indicam tempo de divergência. BEP: subfamílias *Bambusoideae*, *Ehrhartoideae* e *Pooideae*. Adaptado de SETTA *et al.*, 2014.

De todas as espécies vegetais conhecidas, a eudicotiledônea *Arabidopsis thaliana*, da família *Brassicaceae*, é de longe aquela com maior quantidade de estudos. Foi originalmente adotada como organismo modelo devido à sua utilidade em experimentos genéticos, possuindo características como curto período de geração, pequeno tamanho (que permite o cultivo em locais reduzidos) e a grande quantidade de sementes que pode ser obtida por autopolinização (MEINKE *et al.*, 1998). O sequenciamento do genoma de *Arabidopsis* (*Arabidopsis Genome Initiative*, 2000) criou uma ferramenta de valor inestimável para a identificação eficiente e análises funcionais sistemáticas dos genes envolvidos na síntese da parede celular vegetal. Apesar da complexidade da estrutura e da biossíntese da parede, estudos genéticos em *Arabidopsis* levaram à identificação de muitos genes envolvidos na sua biogênese. Esses estudos tem um grande impacto no entendimento da parede, assim como da função de cada polímero dentro da parede (SCHELLER; ULVSKOV 2010).

A *Arabidopsis thaliana* tem um papel de espécie pioneira como modelo genético para estudos de crescimento e desenvolvimento de plantas. Entretanto, as gramíneas

divergiram das eudicotiledôneas na história evolutiva das plantas há cerca de 130-150 milhões de anos (VANNESTE *et al.*, 2014), o que evidencia a necessidade de modelos de estudos específicos para gramíneas, como brachypodium, arroz, sorgo e milho (CARPITA e MCCANN, 2008; KELLOGG, 2015, EDISON *et al.*, 2016) em especial no que diz respeito à parede celular.

1.2 Parede Vegetal

As plantas capturam a energia da luz solar, convertendo dióxido de carbono e água em carboidratos através do processo de fotossíntese, provendo então, direta ou indiretamente, a maior parte da energia utilizada pela vida na Terra. Através da fotossíntese, grande quantidade de energia é armazenada nos polissacarídeos estruturais que compõem a parede celular, o principal componente da biomassa (WEI *et al.*, 2009).

A parede celular é considerada uma valiosa matéria-prima para a produção de biocombustíveis de segunda geração e outros compostos químicos. Esses materiais lignocelulósicos podem ter origem na agricultura ou como resíduos industriais, como a palha de milho, bagaço de cana-de-açúcar, ou ainda de plantas cultivadas especificamente para a produção de grandes quantidades de lignocelulósicos (MIEDES *et al.*, 2014).

A biomassa lignocelulósica pode ser convertida em açúcares simples, que são então fermentados a combustíveis líquidos através de processos químicos e biológicos. Entretanto, esse processo ainda é ineficiente e requer altos investimentos. O principal gargalo na tecnologia de conversão é a recalcitrância, resistência natural da parede celular à desconstrução biótica e enzimática até açúcares simples. Isso pode ser mitigado aumentando o rendimento de açúcar no pré-tratamento ou melhorando a atividade das enzimas e/ou seu custo (MIEDES *et al.*, 2014), além da melhoria do substrato (planta).

A otimização do uso da matéria-prima para biocombustíveis pode ser dividida em dois aspectos: a quantidade e a qualidade de biomassa. A quantidade é expressa em toneladas de material por área plantada, e a qualidade é avaliada em termos de litros de biocombustível produzido por unidade da biomassa. No contexto de produção de etanol, uma biomassa de alta qualidade é aquela que pode ser convertida facilmente e com baixo custo a combustíveis líquidos. Altas quantidades de polissacarídeos são importantes, mas a qualidade também o é, pois açúcares como a hemicelulose podem estar ligados à lignina e, assim, inacessíveis ao processamento (Figura 2). O conteúdo de lignina e o tipo de ligação

entre ela e a hemicelulose e celulose são fatores determinantes dessa qualidade da biomassa. A recalcitrância da parede celular aumenta o custo da conversão da celulose à glicose (BYRT *et al.*, 2011). Novos conceitos vêm sendo propostos a fim de alcançar o objetivo geral de reduzir os custos do processo, incluindo-se a modificação genética da parede celular de linhagens energéticas a fim de deixá-las mais facilmente degradáveis (WEI *et al.*, 2009; CHEN *et al.*, 2007; STICKLEN, 2008; WENG *et al.*, 2008).

Figura 2 - Interação entre os componentes da parede celular: polissacarídeos e lignina.

Representação esquemática das possíveis ligações, sem levar em conta suas reais frequências em materiais biológicos. Ligação direta de éster (1); ligação direta de éter (2); éster de ácido hidroxicinâmico (3); éter de ácido hidroxicinâmico (4); ponte de ácido ferúlico (5); ponte diéster de ácido ferúlico (6) e; ponte diéster-éter de ácido ferúlico (7). (BUCHANAN *et al.*, 2000)

Morfologicamente, a parede celular é dividida em três regiões distintas, a lamela média, parede primária e parede secundária (Figura 3). A lamela média é a região de interseção entre duas células adjacentes. Esta camada intercelular é rica em pectinas (BUCHANAN *et al.*, 2000).

Figura 3 - Diagrama da organização da parede celular vegetal.

Três camadas distintas (S1, S2 e S3) são formadas internamente à parede primária. (TAIZ e ZEIGER, 2004)

Quando o crescimento celular é finalizado, inicia-se o processo de diferenciação celular com início da formação da parede secundária. O início da formação da parede secundária é marcado pela redução da síntese de pectina e um grande aumento na síntese de celulose, hemiceluloses e lignina (BOUDET *et al.*, 2003).

A lignificação é a fase final da diferenciação celular, onde a lignina depositada estabelece ligações covalentes com a matriz de carboidratos, formando o chamado complexo lignina-carboidratos. O processo de diferenciação destas células resulta em uma intensa lignificação e na morte celular programada. A parede celular secundária aumenta a rigidez da célula e reduz a suscetibilidade da planta aos estresses bióticos e abióticos (FENGEL; WEGENER, 1989).

A parede celular secundária contém tipicamente três camadas distintas, S1, S2 e S3, as quais são diferenciadas em espessura, composição e orientação das microfibrilas de celulose. A camada exterior S1 é a mais fina e representa apenas 5-10% da espessura total da parede celular. Nesta camada, as microfibrilas de celulose formam um ângulo de 60-80° em relação ao eixo da célula. A camada intermediária S2 é a mais espessa e representa 75-85% da espessura total da parede celular, e a orientação das microfibrilas forma um ângulo

de 5-30° em relação ao eixo da célula. As características relacionadas à espessura e à orientação das microfibrilas influenciam diretamente as propriedades físicas e mecânicas da célula. Já a camada S3 é relativamente fina, e forma um ângulo de 60-90° em relação ao eixo da célula (EMONS *et al.*, 2000).

A estrutura e composição da parede celular variam entre as espécies de plantas, diferentes tecidos e células, entre as paredes primária e secundária da mesma célula, e até mesmo entre as diferentes camadas da mesma parede celular. A complexidade química e estrutural da parede celular permite sua variedade de funções biológicas, incluindo suporte mecânico, proteção contra patógenos e transporte de água e solutos no nível celular e de organismo, e ao mesmo tempo é um fator causador da recalcitrância na catálise química e biológica para produção de biocombustíveis. A parede celular secundária é composta de polissacarídeos, lignina, proteínas e compostos cerosos; esses compostos são independentes, porém interativos entre si (HIMMEL *et al.*, 2007; BUCHANAN *et al.*, 2000). A quantidade relativa de cada composto varia entre as diferentes paredes. A celulose e a hemicelulose são os dois principais polissacarídeos visados como fonte de açúcar para produção de biocombustíveis (WEI *et al.*, 2009).

Apesar de a biomassa ser frequentemente considerada como tendo uma composição uniforme, há uma substancial variação na sua composição que tem origem em dois aspectos. Primeiro, diferentes espécies tem diferentes proporções de celulose, hemicelulose e lignina (Tabela 1) e, segundo, diferenças no tipo de hemicelulose e na proporções dos componentes da lignina (PAULY *et al.*, 2008; VANHOLME *et al.*, 2008).

Tabela 1 - Composição da parede celular vegetal em diferentes espécies.

Espécie	% de celulose	% de hemicelulose	% de lignina
Milho (palha)	35%	28%	10%
Miscanthus	57%	16%	10%
Poplar	40%	14%	20%
Cana-de-açúcar	38-43%	25-32%	17-24%
Sorgo	26%	20%	7%
Switchgrass	31%	36%	6%
Salgueiro	56%	14%	19%

(Adaptado de Byrt *et al.*, 2010 e Masarin *et al.*, 2011)

A celulose é um agrupamento de cadeias lineares poli β -1,4-glucana unidas por ligações de hidrogênio inter-cadeia (HAIGLER, 1997; TAYLOR *et al.*, 1992; DING *et al.*, 2006; ATALLA *et al.*, 2008). As propriedades biofísicas da celulose são baseadas quase

totalmente na geometria com a qual essas cadeias estão organizadas e no diâmetro da fibrila, e podem variar conforme o solvente no qual são estudadas (ATALLA *et al.*, 2008; WEI *et al.*, 2009).

As hemiceluloses e pectinas são polissacarídeos de cadeia ramificada. Esses polissacarídeos não celulósicos são compostos de mais de vinte monossacarídeos diferentes em diversas combinações. Estudos demonstram que as hemiceluloses e as pectinas possuem ligações cruzadas com as microfibrilas de celulose e formam ligações covalentes com outros polímeros da parede, como a lignina (WEI *et al.*, 2009).

1.3 Lignina

A lignina é o segundo biopolímero mais abundante, atrás apenas da celulose (WENG *et al.*, 2008; DAVIN *et al.*, 2005), e constitui um enorme acúmulo de carbono na biosfera, contando com aproximadamente 30% dos mais de $1,4 \times 10^{12}$ kg de carbono sequestrados da atmosfera anualmente (BATTLE *et al.*, 2000). Como um dos principais polímeros da parede celular, a lignina tem um impacto direto nas características da biomassa, incluindo a facilidade com que a lignina indesejável é removida das fibras de celulose durante o processo de polpação (HUMPHREYS; CHAPPLE, 2002), além de impactar na conversão de materiais lignocelulósicos em biocombustíveis (ZENG *et al.*, 2014). Ainda, o suporte mecânico que provem da lignina impede o tombamento da planta, um aspecto importante em diversas culturas agrônômicas. A lignina reduz a digestibilidade de espécies forrageiras como alfalfa (*Medicago sativa*), o que se converte em impacto na indústria de pastagem por diminuir seu rendimento na nutrição animal (MOORE *et al.*, 2001). Finalmente, as muitas funções da lignina e seus co-produtos na resistência a estresse biótico e abiótico fazem com que a via dos fenilpropanóides seja vital para a sobrevivência da planta (ZENG *et al.*, 2014).

A via dos fenilpropanóides é responsável pela biossíntese de uma variedade de produtos que inclui a lignina, flavonóides e ácidos hidroxicinâmicos conjugados. Muitos intermediários e produtos finais dessa via têm papéis importantes na planta como fitoalexinas, antiherbívoros, antioxidantes, proteção contra UV, pigmentos e compostos aromáticos (DIXON, 2001; DAVIN *et al.*, 1992; HAHLBROK *et al.*, 1989). A maioria do carbono nessa via, entretanto, é direcionada à síntese de lignina, um dos componentes principais da parede celular vegetal (HUMPHREYS; CHAPPLE 2002).

A lignina é um heteropolímero racêmico aromático complexo, derivado basicamente de três principais álcoois hidroxicinâmicos que diferem entre si pelo seu grau de metoxilação, o *p*-coumaril, coniferil e o sinapil (FREUDENBERG; NEISH, 1968; KUMAR *et al.*, 2016) (Figura 4). Esses monolignóis produzem, respectivamente, as unidades fenilpropanóides *p*-hidroxifenil (H), guaiacil (G) e siringil (S) quando incorporados ao polímero de lignina. A quantidade e composição da lignina variam entre espécies, tipos celulares e camadas da parede de uma célula individual, e são influenciadas por fatores do desenvolvimento e do ambiente no qual o vegetal se encontra (CAMPBELL *et al.*, 1996; BHUIYAN, 2009). Apesar de existir exceções, a lignina das angiospermas eudicotiledôneas (madeiras duras, ou *hardwood*, como o Eucalipto) consiste principalmente de unidades G e S e traços de unidades H, enquanto que a de gimnospermas (madeiras macias, ou *softwood*, como a Araucária) é composta de unidades G e baixas quantidades de unidades H. A lignina de gramíneas (monocotiledôneas) incorpora unidades G e S em níveis semelhantes, e mais unidades H do que as eudicotiledôneas (BOERJAN *et al.*, 2003). Unidades H, derivadas da incorporação de álcool *p*-coumaril, não devem ser confundidas com ésteres *p*-coumarato, os quais aparecem como grupos substituintes acetilados da lignina ou como ésteres ligados à hemicelulose e à lignina.

Figura 4 - Principais monolignóis.

H: álcool *p*-cumárico; G: álcool coniferil; S: álcool sinapil. Fonte: do autor.

A deposição de lignina é o processo pelo qual os monolignóis são ligados via acoplamento radical de cada monômero ao polímero crescente, predominantemente na extremidade fenólica (BOERJAN *et al.*, 2003; SMITH *et al.*, 2015). A união de dois monolignóis é um evento menos frequente, e substancialmente diferente da formação de lignina (ADLER, 1977). Muitas plantas apresentam lignina derivada de outros monômeros, e todas as ligninas de diferentes vegetais contém traços de unidades fruto de biossíntese aparentemente incompleta ou de outras reações alternativas durante a biossíntese. Muitas dessas unidades vêm sendo identificadas em quantidades elevadas em plantas

geneticamente modificadas e mutantes, produtos de plantas com perturbações na via de biossíntese de lignina (RALPH *et al.*, 2001).

A lignificação é um dos estágios finais da diferenciação celular do xilema e ocorre predominantemente no durante o espessamento da parede secundária (DONALDSON, 2001). Em geral, a parede secundária consiste de três camadas: a externa (S1), média (S2) e a interna (S3). A deposição de lignina ocorre em diferentes fases, cada qual precedida pelo acúmulo de carboidratos, e tem início nos cantos das células, regiões onde a formação da lamela média e de S1 se inicia (Figura 3). Quando a formação da matriz de polissacarídeos na camada S2 se completa, a lignificação ocorre na parede secundária. Geralmente, a concentração de lignina é maior na lamela média e nos cantos das células do que na camada S2 (BAUCHER *et al.*, 1998; DONALDSON, 2001; SAKA *et al.*, 1985). Entretanto, como a camada S2 ocupa a maior porção da parede, ela abriga a maior quantidade de lignina. Paredes de vasos geralmente possuem maior teor de lignina do que as paredes de fibras. Condições ambientais podem, também, influenciar a quantidade e a composição da lignina (DONALDSON, 2001).

Estudos demonstram que os três monolignóis são incorporados em diferentes estágios na formação da parede. Tipicamente, unidades H são depositadas inicialmente, seguidas pelas unidades G e então, em angiospermas, as unidades S (DONALDSON, 2001; TERASHIMA *et al.*, 1993). A lignina de vasos é geralmente rica em unidades G, enquanto a lignina de fibras é rica em unidades S (SAKA *et al.*, 1985). Em gramíneas (monocotiledôneas), a lignina também incorpora grandes quantidades de ésteres hidroxycinamatos (ferulato e *p*-coumarato) (CASS *et al.*, 2015).

A natureza química da matriz de carboidratos e a orientação das microfibrilas de celulose influenciam a deposição da lignina. Na lamela média e na parede primária, a lignina forma estruturas esféricas, enquanto que na parede secundária ela forma lamelas que seguem a orientação das microfibrilas (ATALLA *et al.*, 1985; DONALDSON, 2001; ROUSSEL *et al.*, 1995). Durante a deposição, a lignina pode formar ligações químicas com a hemicelulose da parede e eliminar água gradualmente, formando um ambiente hidrofóbico.

C4H ter sido demonstrado (RASMUSSEN *et al.*, 1999; ACHNINE *et al.*, 2004), permanece incerto se as outras enzimas da via também fazem parte de um complexo metabólico no retículo endoplasmático (VANHOLME *et al.*, 2010).

Para a maioria das enzimas descritas acima, existem múltiplas isoformas que são diferentemente expressas durante o desenvolvimento e em resposta a estímulos ambientais (CHEN *et al.*, 2000; HU *et al.*, 1998; LAUVERGEAT *et al.*, 2001; LINDERMAYR *et al.*, 2002) e devem possuir diferentes cinéticas e substratos preferenciais (EHLTING *et al.*, 1999; HARDING *et al.*, 2002; ZUBIETA *et al.*, 2002). É de se esperar, então, que certos passos da via sejam favorecidos cineticamente em determinadas células ou condições ambientais, permitindo um metabolismo flexível. O trabalho de Li e colaboradores (2015) demonstrou a existência de quatro isoformas de 4CL em *Arabidopsis* com funções distintas, apesar de certa sobreposição funcional. Outro nível de complexidade vem do fato de que intermediários da via podem afetar a síntese ou a atividade de certas enzimas da própria via. As concentrações de fenilalanina também têm grandes efeitos na regulação da via, por se tratar do composto inicial.

Além de uma flexibilidade no metabolismo e no conteúdo total de lignina, é observada também uma flexibilidade nas proporções entre as unidades H, G e S (VANHOLME *et al.*, 2010). Mudanças na composição de lignina, isto é, na proporção dos resíduos H, G e S, podem ser realizadas por meio da alteração na síntese de cada monômero, reduzindo sua disponibilidade no momento da polimerização (VANHOLME *et al.*, 2010; VANHOLME *et al.*, 2012). Estudos demonstram que uma maior proporção S/G melhora a degradação química da parede celular de monocotiledôneas e eudicotiledônias, otimizando a produção de bioetanol celulósico (FORNALÉ *et al.* 2012; MANSFIELD *et al.* 2012). Ainda, a razão S/G afeta o rendimento da sacarificação quando o material é submetido à hidrólise enzimática após pré-tratamento (LI *et al.* 2010; VAN ACKER *et al.* 2013; VERMERRIS *et al.*, 2015). Porém, essa razão provavelmente não afeta digestibilidade da parede celular se não for realizado pré-tratamento (GRABBER *et al.*, 1997; GRABBER, 2005; REDDY *et al.*, 2005; LI *et al.*, 2010; PARIJS, 2015).

Ensaio *in vitro* de enzimas individuais da via e experimentos de radiomarcagem são instrutivos em dissecar a rota de biossíntese dos monolignóis, mas são insuficientes para se compreender a complexidade da via *in vivo*. Estudos com plantas transgênicas nos permitem visualizar os efeitos das alterações na quantidade de lignina, sua composição e

estrutura primária, além dos efeitos fenotípicos de expressão alterada de um gene específico. Assim, pode-se obter embasamento dos aspectos regulatórios da via na estrutura e nos papéis biológicos da lignina (BOERJAN *et al.*, 2003).

Os impactos causados pela lignina tem amplo alcance na agricultura, indústria e no meio ambiente, fazendo com que a via dos fenilpropanóides seja de enorme importância no contexto da bioquímica da planta (HUMPHREYS; CHAPPLE, 2002). A biossíntese de lignina é um campo de pesquisa ativo, em parte por sua relevância econômica. Para a produção de papel de alta qualidade, a lignina precisa ser removida da polpa por meios custosos e ambientalmente indesejáveis, necessitando de alto gasto energético e químico. Ainda, a eficiência da conversão de biomassa em biocombustíveis depende fortemente do melhoramento das tecnologias de pré-tratamento e, especificamente, na efetividade na modificação da lignina (BONAWITZ; CHAPPLE, 2010).

O aumento da quantidade de resíduos H em virtude da diminuição dos resíduos S e G pode ser obtida através da repressão de *C3H* ou *HCT* (CHEN *et al.*, 2007). O mutante de *Arabidopsis ref8-1*, que possui uma baixa atividade de *C3H*, resulta em uma lignina quase que exclusivamente composta de resíduos H, e apresenta um fenótipo anão, que pode indicar a inabilidade da lignina puramente H de dar suporte ao crescimento e desenvolvimento da planta (VERMERRIS *et al.*, 2015). O silenciamento de *HCT* de alfalfa (*Medicago sativa*) reduz o conteúdo de lignina e melhora a eficiência da sacarificação para produção de etanol (GALLEGO-GIRALDO *et al.*, 2010).

Em outro trabalho, em *Eucalyptus*, o silenciamento de genes *C3H* e *C4H* provocaram redução no conteúdo de lignina e aumento na liberação de açúcares (SYKES *et al.*, 2015). Ainda, Shafrin e colaboradores (2015) realizaram o silenciamento de *C3H* e *F5H* em juta (*Corchorus olitorius*) e observaram a redução no conteúdo de lignina, sem perda da capacidade de defesa da parede celular. A lignina do mutante *fah-1* de *Arabidopsis*, que possui atividade reduzida de *F5H*, caracteriza-se pela ausência de unidades S (BOERJAN *et al.*, 2003).

A repressão de *PAL* e *C4H* em tabaco reduz fortemente o conteúdo de lignina (BATE *et al.*, 1994; ELKIND *et al.*, 1990; SEWALT *et al.*, 1997), porém com efeitos colaterais na composição de lignina. A baixa expressão da *PAL* leva a uma redução nos níveis de unidades G em brachypodium (CASS *et al.*, 2015), enquanto que a baixa expressão de *C4H* reduz os níveis de unidades S (BOERJAN *et al.*, 2003). Ainda, estudos

demonstram que o silenciamento de *PAL* em *brachypodium* gera um impacto positivo na disponibilidade de carbono para vias de interesse na produção de biocombustíveis, porém causa também uma redução na resistência a doenças (CASS *et al.*, 2015). O trabalho de Shen e colaboradores (2013) realizou o silenciamento da enzima *C4H* de *switchgrass* (*Panicum virgatum*) em 70%, porém sem encontrar efeito no conteúdo de lignina nem tampouco na sua composição.

A baixa expressão de *4CL* leva à redução do conteúdo de lignina em tabaco (*Nicotiana tabacum*), *Arabidopsis* e aspen (*Populus tremuloides*) (HU *et al.*, 1999; KAJITA *et al.*, 1997; LEE *et al.*, 1997). A redução na *4CL* de *switchgrass* leva a uma diminuição no conteúdo de lignina e aumento na proporção S/G (XU *et al.*, 2011). A diminuição da expressão de *CCoAOMT*, por sua vez, gera redução no conteúdo de lignina e aumento na relação S/G, principalmente devido à diminuição das unidades G. Apesar disso, Shen e colaboradores (2013) não encontraram efeito no silenciamento de *CCoAOMT* em *switchgrass*. Por outro lado, a repressão de *COMT* leva, além da diminuição do conteúdo de lignina, à diminuição proporcional das unidades S (GUO *et al.*, 2001; JOUANIN *et al.*, 2000; LAPIERRE *et al.*, 1999; TORAMAN *et al.*, 2016). Ainda, no estudo realizado por Ho-Yue-Kuang e colaboradores (2015), uma mutação na enzima *COMT* de *brachypodium* mostrou alteração na composição de lignina, com aumento na quantidade relativa de unidades G, porém sem efeito exacerbado no conteúdo total de lignina. Assim, não foi observado nenhum comprometimento do desenvolvimento geral da planta.

O estudo de Srivastava e colaboradores (2015) mostra que, em *Leucaena leucocephala*, a indução de estresse abiótico leva a um aumento na expressão de *CCR* e correspondente aumento no conteúdo de lignina em raízes e caule. Em tabaco, a repressão de *CCR* tem efeitos pronunciados tanto na quantidade quanto na composição da lignina (PIQUEMAL *et al.*, 1998; RALPH *et al.*, 1998). A proporção S/G aumenta devido à redução de unidades G, mas a lignina resultante é mais condensada. A redução da transcrição de *CCR* em *switchgrass* leva a uma diminuição no rendimento na tioacidólise de lignina (SHEN *et al.*, 2013).

Apesar de a CAD catalisar o último passo da via de biossíntese dos monolignóis, o conteúdo de lignina não sofre alteração expressiva em plantas deficientes em CAD (BAUCHER *et al.*, 1999; BAUCHER *et al.*, 1996). Isso pode ser explicado, em partes,

pela possibilidade de incorporação de outras moléculas na parede celular, na condição de limitada disponibilidade de monolignóis (BOERJAN *et al.*, 2003). Entretanto, alterações na função dessa enzima podem levar a reduções no crescimento da planta (ANDERSON *et al.*, 2015). Um estudo com silenciamento de *CAD* em *brachypodium* demonstrou diminuição no conteúdo de lignina S, e consequente redução da relação S/G (TRABUCCO *et al.*, 2013). Mutações na *CAD* de arroz foram relacionadas a uma maior resistência do colmo à quebra (OOKAWA *et al.*, 2014), e linhagens transgênicas de *switchgrass* com redução da expressão de *CAD* resultaram em diminuição do conteúdo e alteração na composição da lignina (FU *et al.*, 2011).

Uma lista das espécies de interesse energético que já foram estudadas com relação à alteração de componentes da via pode ser visualizada na tabela 2.

Tomados em conjunto, os resultados obtidos em análises de transgênicos modificados na via de biossíntese dos monolignóis demonstram que as plantas podem suportar variações no conteúdo e composição da lignina; moléculas diferentes dos três principais monômeros podem ser incorporadas e; a incorporação desses diferentes monômeros resulta em novas estruturas de lignina. Isso mostra como a lignina é um polímero extremamente flexível em sua composição. Além de contribuírem no entendimento da biossíntese e da estrutura de lignina, esses experimentos com transgênicos mostram também como a lignina é importante para a integridade estrutural da parede celular (PINÇON *et al.*, 2001a; PINÇON *et al.*, 2001b).

Tabela 2 - Espécies de interesse energético já estudadas.

Enzima	Espécie	Manipulação da Lignina		Sacarificação por hidrólise enzimática			Referências
		Método de manipulação	Conteúdo de Lignina	Proporção S : G	Pré-tratamento	Eficiência de sacarificação	
4CL	Populus	AS	↓ 10%	↓	HW	NM	Voelker et al. (2010)
	Switchgrass	RNAi	↓ 17%–32%	↓	DA	↑ 57%	Xu et al. (2011)
C4H	Alfafa	AS	↓	↓	Nenhum	↑	Chen and Dixon (2007)
HCT	Alfafa	AS	↓	↑	DA	↑	Chen and Dixon (2007)
C3H	Alfafa	AS	↓	↑	DA	↑	Chen and Dixon (2007)
	Alfafa	AS	↓	↑	Nenhum	↑	Chen and Dixon (2007)
CCoAOMT	Alfafa	AS	↓	↑	DA	NM	Chen and Dixon (2007)
CCR	Populus	AS	↓ 5%–24%	ND	DA	↑ 14%–139%	Van Acker et al. (2014)
	Populus	AS	↓ 5%–24%	ND	DA	↑ 19%–95%	Van Acker et al. (2014)
F5H	Alfafa	AS	NM	↓	A	↑ 11%–78%	Chen and Dixon (2007)
COMT	Alfafa	AS	↓	↓	Nenhum	NM	Chen and Dixon (2007)
	Alfafa	AS	↓	↓	DA	NM	Chen and Dixon (2007)
Switchgrass (estufa)	Switchgrass (estufa)	RNAi	↓ 6%–14%	↓	Nenhum	↑	Fu et al. (2011a)
	Switchgrass (campo)	RNAi	↓ 8%–15%	↓	DA	↑ 29%–38%	Baxter et al. (2014)
	Cana-de-açúcar (estufa)	RNAi	↓ 4%–14%	↓	HW	↑ 17%–22%	Jung et al. (2012)
	Cana-de-açúcar (campo)	RNAi	↓ 6%–12%	↓	Nenhum	↑ 6%–34%	Jung et al. (2013)
CAD	Switchgrass	RNAi	↓ 14%–22%	↓	DA	↑ 19%–32%	Fu et al. (2011b)
	Switchgrass	RNAi	↓	ND	Nenhum	↑ 19%–89%	Saathoff et al. (2011)
	Milho	RNAi	NM em vasos	↓	A	↑ 2%–11%	Fornale et al. (2012)
	Milho	RNAi	NM em vasos	↓	Nenhum	↑ 25% em vasos	Fornale et al. (2012)
↓ 6.4% em midribs							↑ 16% em midribs

Valores representam o percentual em relação ao controle. Setas representam aumento ou diminuição. AS, sub-expressão mediada por antissenso; RNAi, RNA de interferência; OE, super-expressão; ND: não disponível; NM: nenhuma mudança; HW, água quente; DA, ácido diluído; and A, alcalino. Adaptado de Poovaiah et al (2014).

1.5 Cana-de-açúcar: uso como matéria-prima, genética e variedades

A cana-de-açúcar (*Saccharum* spp.) é uma gramínea perene tropical que é cultivada pelo seu conteúdo de sacarose. Em determinados países, como os Estados Unidos da América, é cultivada para a produção de açúcar, porém outros países como o Brasil a utilizam, também, para a produção de bioetanol (LAKSHMANAN *et al.*, 2005). A cana-de-açúcar é uma das mais importantes culturas do mundo, sendo a fonte primária de açúcar para alimentos e bebidas e para a produção de biocombustíveis. As dificuldades e limitações associadas ao melhoramento tradicional de cana-de-açúcar, além do grande potencial do uso de engenharia genética, tornam a biotecnologia uma abordagem interessante para o melhoramento de aspectos relevantes da cana-de-açúcar, como conteúdo de açúcar e tolerância a estresse biótico e abiótico (LAKSHMANAN *et al.*, 2005). Ainda, a biotecnologia pode ser utilizada para diversificação da indústria de cana-de-açúcar, criando linhagens que produzam produtos alternativos de alto valor agregados, como compostos farmacêuticos, antibióticos, produtos industriais e açúcares alternativos (WANG *et al.*, 2005; BASNAYAKE *et al.*, 2012; HARRISON *et al.*, 2011; PETRASOVITS *et al.*, 2012; KINKEMA *et al.*, 2014).

Além de ser uma importante fonte de alimento e energia, há várias outras razões pelas quais a cana-de-açúcar é uma candidata ao melhoramento por engenharia genética. Primeiro, o melhoramento convencional por cruzamento com clones de elite é muito difícil devido ao genoma poliploide-aneuplóide, a baixa fertilidade, e o longo período necessário para se criar novos cultivares (de dez a quinze anos). Nesse contexto, a engenharia genética é uma ferramenta valiosa para se introduzir características de interesse comercial no germoplasma. Em segundo lugar, já existem sistemas práticos de transformação em cana-de-açúcar (ARENCIBIA *et al.*, 1998; ENRIQUEZ-OBREGON *et al.*, 1998; MANICKAVASAGAM *et al.*, 2004) e linhagens transgênicas podem ser mantidas quase que indefinidamente por propagação vegetativa (KEATING *et al.*, 1997).

Em todo o mundo, a produção de cana-de-açúcar tem crescido progressivamente a níveis extraordinários no último século (ROACH, 1989). O aumento na produtividade se deve ao desenvolvimento e difusão de cultivares melhorados com aumento na resistência a doenças e pestes, melhor manejo da água, nutrientes e recursos, e a disponibilidade de fertilizantes e pesticidas relativamente baratos. Sustentar esse ritmo de melhorias na

produtividade, garantindo mínimo impacto ambiental, será um dos grandes desafios para se manter a indústria de cana-de-açúcar lucrativa no futuro (LAKSHMANAN *et al.*, 2005).

Até o final do século XIX, a maioria dos cultivares de cana-de-açúcar eram clones de *Saccharum officinarum* com alto teor de sacarose que continham $2n=80$ cromossomos. Um grande salto ocorreu no melhoramento ao se desenvolver o primeiro híbrido entre *S. officinarum* e o selvagem *S. spontaneum*. Uma série de cruzamentos resultou em cultivares com alto rendimento e resistência a doenças. Os cultivares modernos foram desenvolvidos a partir desses híbridos com conteúdo diploide e possuem entre $2n=100$ e $2n=130$ cromossomos. Estima-se que apenas poucos clones de *S. officinarum* e *S. spontaneum* tenham sido envolvidos no desenvolvimento desses híbridos (ROACH, 1989). Os programas modernos de melhoramento de cana-de-açúcar se baseiam extensivamente no cruzamento de cultivares de elite derivados desses primeiros híbridos (LAKSHMANAN *et al.*, 2005).

Estudos recentes com RNA ribossomal sugerem que *S. officinarum* tem um genoma octaplóide (D'HONT *et al.*, 1998), enquanto que *S. spontaneum* teria um valor de ploidia variando entre 5 e 16 (D'HONT *et al.*, 1998; HA *et al.*, 1999). Esses estudos estabeleceram a coexistência de duas organizações cromossômicas distintas nos cultivares modernos. Utilizando a técnica de hibridização genômica *in situ* (GISH, na sigla em inglês), foi demonstrado que os cultivares modernos contém cerca de 15-20% dos seus cromossomos proveniente de *S. spontaneum* e menos de 5% de cromossomos recombinantes (D'HONT *et al.*, 1996; CUADRADO *et al.*, 2004). A alta ploidia e a complexidade do genoma geram desafios para a expressão de transgenes e para o desenvolvimento de marcadores moleculares (LAKSHMANAN *et al.*, 2005).

Durante o processamento da cana-de-açúcar, grandes quantidades de bagaço de cana-de-açúcar são geradas após a maceração desta para retirada do caldo dos colmos. A quantidade de bagaço disponível por tonelada de cana-de-açúcar varia entre 250 e 300 kg (REIN, 2007). Em geral, o bagaço é queimado para produzir vapor e eletricidade como parte do esquema de co-geração nas usinas. O bagaço residual é descartado como dejetos industriais de baixo valor. O conteúdo dos componentes no bagaço varia, sendo em média 41,6% celulose, 25,1% hemicelulose e 20,3% lignina (KIM; DAY, 2011).

As tecnologias de uso da cana-de-açúcar como um sistema biológico para produção em larga escala de produtos de alto valor agregado vêm se desenvolvendo rapidamente. Apesar da dificuldade em se realizar transformações gênicas em cana-de-açúcar, devido à instabilidade do transgene (ARRUDA, 2012), progressos significativos veem sendo reportados, como o estudo de Jung e colaboradores (2012) onde uma redução no conteúdo de lignina e na relação S/G foi observada após silenciamento da enzima COMT via RNA de interferência.

Produtos alternativos à sacarose podem gerar não apenas um valor elevado, mas permitir também que a indústria da cana-de-açúcar seja mais sustentável e competitiva. A cana-de-açúcar possui todas as características necessárias para ser uma biofábrica natural: cresce rapidamente, fixa carbono eficientemente, produz muita biomassa, possui um sistema de reserva bem desenvolvido com um grande *pool* de açúcares hexoses, e é cultivado em diversas partes do mundo (LAKSHMANAN *et al.*, 2005).

As informações sobre o metabolismo de lignina em cana-de-açúcar estão começando a ser desvendados (RAMOS *et al.*, 2001; PAPINI-TERZI *et al.*, 2005; BOTTCHEER *et al.*, 2013; SANTOS *et al.*, 2015; VICENTINI *et al.*, 2015). Vinte e três genes envolvidos na via de biossíntese de lignina foram identificados em estudos de transcriptoma de larga escala por PAPINI-TERZI e colaboradores (2005), que identificaram quatro *PALs*, cinco *C3H/F5H/C4Hs*, três *4CLs*, um *HCT*, três *CCoAOMTs*, dois *CCRs*, um *COMT* e quatro *CADs*.

Bottcher e colaboradores (2013) realizaram um estudo analisando a deposição de lignina e a proporção S/G ao longo do desenvolvimento do caule de cana-de-açúcar, seguido de busca no banco SUCEST por sequências e posterior análise de expressão dessas sequências, a fim de encontrar os melhores candidatos à via de biossíntese de lignina em plantas. Um total de 28 unigenes foram identificados, entre eles quatro *PALs*, duas *C4Hs*, três *4CLs*, uma *HCT*, duas *C3Hs*, três *CCoAOMTs*, duas *CCRs*, uma *COMT*, uma *F5H* e quatro *CADs*. Para enzimas PAL, C4H, CCoAOMT, CCR, COMT e CAD, mais de uma sequência foi utilizada na construção de um único unigene, e por isso o total de sequências do SUCEST listadas como candidatas somam 33 no total.

Baseando-se nas sequências encontradas no trabalho anterior (BOTTCHEER *et al.*, 2013), Santos e colaboradores (2015) avaliaram efeitos do estresse hídrico na expressão dos genes da via de biossíntese de lignina, e concluíram que um aumento de quase 60% no

conteúdo de lignina na medula de internos jovens está relacionado ao aumento de *4CL* e *COMT*, o que sugere que esses genes podem ter papel essencial no controle da deposição de lignina em cana-de-açúcar.

Vicentini e colaboradores (2015), utilizando a tecnologia de RNA-Seq e montagem *de novo*, construíram um banco de transcritos dos quais 21 foram identificados como envolvidos na biossíntese de lignina, sendo duas *PALs*, três *C4Hs*, duas *4CLs*, uma *HCT*, duas *C3Hs*, três *CCoAOMTs*, uma *CCR*, duas *COMTs*, uma *F5H*, e quatro *CADs*. Esse trabalho pôde ainda analisar níveis de expressão diferencial e comparar dois híbridos contrastantes no conteúdo de lignina, buscando correlacionar essas informações.

1.6 Bioinformática

A genômica, análise da sequência completa de DNA de um organismo, tem sido de grande importância em estudos biológicos. A sequência do genoma de um organismo é de fundamental importância para a compreensão das funções individuais dos genes e das redes de interação entre eles, pois permite o conhecimento de relações evolutivas e pode revelar mecanismos regulatórios até então desconhecidos (BEVAN; UAUY, 2013).

Desde a liberação do genoma de *Arabidopsis* em 2000 (*Arabidopsis Genome Initiative*), um número crescente de espécies de plantas vem tendo seu genoma sequenciado, e isso revoluciona nosso entendimento da biologia vegetal e facilita o melhoramento das espécies de interesse (VARSHNEY, 2012; BOLGER, 2014, BEVAN; UAUY, 2013). Novas metodologias biológicas vêm se desenvolvendo devido às tecnologias ômicas, primeiramente a análise de DNA (genômica), seguido pelo RNA (transcriptômica), proteínas (proteômica) e pequenos metabólitos (metabolômica) (WANG *et al.*, 2015).

Os recentes avanços no sequenciamento de genomas através, por exemplo, das tecnologias de sequenciamento de nova geração, permitem o estudo de espécies vegetais não consideradas modelos, como a busca de genes de interesse agrônomico. A identificação desses genes em diferentes espécies permite um entendimento de como algumas características vegetais de interesse são reguladas, conhecimento esse que pode ser utilizado diretamente no melhoramento vegetal (EDWARDS; BATLEY, 2010).

Os avanços recentes das tecnologias de sequenciamento permitiram o sequenciamento de transcriptomas pelo método de RNA-Seq, que permite tanto mapear quanto quantificar o RNA mensageiro de uma amostra (YANG; KIM, 2015; WANG *et al.*, 2009). Em geral, uma população de RNA (total ou fracionado) é convertido em uma biblioteca de cDNA com adaptadores ligados às extremidades das moléculas. Cada molécula é então, com ou sem amplificação prévia, sequenciada a fim de se obter pequenas sequências das extremidades. Em seguida, esses fragmentos são alinhados a um genoma de referência ou utilizados para a montagem de um novo mapa de transcritos, podendo incluir dados de nível de expressão para cada gene (WANG *et al.*, 2009). Ainda, RNA-Seq pode fornecer informações sobre isoformas de mRNA provenientes de *splicing* alternativos (GRIFFITH *et al.*, 2010) e expressão alelo-específica (HAN *et al.*, 2015).

A identificação de regiões codificantes de proteínas em sequências de nucleotídeos permanece um desafio na interpretação de genomas (STATES; GISH, 1994). A descoberta de homologia entre uma determinada sequência de ácido nucleico e/ou uma proteína é frequentemente o primeiro indício sobre a função de um gene recém-sequenciado. Como os bancos de dados de DNA e de aminoácidos estão em constante expansão, eles se tornam cada vez mais úteis na análise de novas sequências, pois é cada vez maior a chance de se encontrar alguma homologia (ALTSCHUL *et al.*, 1990). Existe um grande número de *softwares* disponíveis para a busca em bases de dados, porém todos utilizam algum método de análise de similaridade entre sequências para distinguir relações de significado biológico daquelas similaridades ao acaso (NEEDLEMAN; WUNSH 1970; SELLERS, 1974; SANKOFF; KRUSKAL, 1983; WATERMAN, 1984).

O BLAST (*Basic Local Alignment Search Tool*) é um programa de busca de similaridade de sequências que pode ser usado via interface na web, ou como ferramenta individual. Existem vários tipos específicos de Blast, que comparam aminoácidos e nucleotídeos com o banco de dados. É um programa heurístico que busca pequenas regiões de similaridade entre duas sequências e tenta iniciar um alinhamento a partir desses “*hot spots*”. Além do alinhamento em si, o Blast oferece dados estatísticos para ajudar a interpretação do significado biológico desse alinhamento (MCGINNIS; MADDEN, 2004).

O HMMER é outro *software* de busca de similaridade, que utiliza o método probabilístico das cadeias ocultas de Markov (*Hidden Markov Models*, HMM). Pode ser utilizado na busca em um banco de dados a partir de uma única sequência proteica, um

grupo de sequências, ou ainda um perfil HMM, que pode ser criado pelo próprio programa. Inclui alguns algoritmos de busca específicos, como o *JackHmmer* e o *HmmerSearch* (FINN *et al.*, 2011).

1.7 Justificativa

Apesar de todo o interesse econômico relacionado a cana-de-açúcar, sua ploidia (8 a 12n) e grande genoma (10Gb) dificultam o sequenciamento para posterior montagem das sequências gênicas. Mesmo assim, desde 2003 diversos esforços no contexto da transcriptômica vêm sendo realizados, inicialmente com o SUCET e atualmente com o RNA-Seq, na tentativa de identificar sequências expressas que facilitem a identificação de genes de interesse, como os genes da via de biossíntese de lignina. Até o momento, cerca de 25 genes homólogos aos genes da via de biossíntese de lignina foram identificados em cana-de-açúcar (Anexo A). Entretanto, apesar deste avanço, os diversos trabalhos não exploram a identificação destes genes dentro de suas respectivas famílias gênicas em abordagem filogenômica possibilitando a identificação dos genes ortólogos e parálogos, bem como identificar e delimitar o clado *bona fide* dos genes envolvidos na via de biossíntese de lignina. Adicionalmente, apesar do uso da tecnologia do RNA-Seq a partir de distintas cultivares de cana-de-açúcar e em diferentes órgãos e tecidos, ainda não houve uma identificação e análise comparativa destas sequências em diferentes órgãos de cana-de-açúcar possibilitando encontrar sequências completas e debater sobre a natureza das sequências expressas.

2 OBJETIVOS

2.1 Objetivos Gerais

O presente trabalho visa à identificação, anotação e filogenia dos genes das famílias gênicas envolvidas na biossíntese de lignina em cana-de-açúcar.

2.2 Objetivos específicos

Identificar e anotar todos os membros das famílias gênicas AMP_binding (enzima 4CL), p450 (C3H, C4H e F5H), ADH_N (CAD), Epimerase (CCR), Methyltransf_3 (CCoAOMT), Methyltransf_2 (COMT), Transferase (HCT) e Lyase_aromatic (PAL) em *Arabidopsis thaliana*, *Sorghum bicolor* (sorgo), *Zea mays* (milho), *Oryza sativa* (arroz) e *Brachypodium distachyon*;

Identificar os genes possivelmente envolvidos na biossíntese de lignina (*bona fides*) nas espécies supra-citadas;

Identificar e anotar as sequências ortólogas e parálogas possivelmente envolvidas na biossíntese de lignina nos diferentes bancos de transcriptoma de cana-de-açúcar.

3 MATERIAL E MÉTODOS

3.1 Identificação e anotação das sequências pertencentes às famílias gênicas relacionadas à biossíntese de lignina em *Arabidopsis thaliana*, gramíneas e cana-de-açúcar

Com base nas sequências descritas na literatura como envolvidas na biossíntese de lignina em *Arabidopsis thaliana*, At2g37040 (PAL), At2g30490 (C4H), At2g40890 (C3H), At4G36220 (F5H), At1g51680 (4CL), At5g48930 (HCT), At5g54160 (COMT), At1g67980 (CCoAOMT), At1g15950 (CCR) e At3g19450 (CAD), obtidas no banco de dados *TAIR* (*The Arabidopsis Information Resource*, <http://www.arabidopsis.org/index.jsp>) (Lamesch *et al.*, 2011), foram realizadas buscas de homólogos nessa mesma espécie e ortólogos em *Sorghum bicolor* (sorgo) (v3.1, *Phytozome* 10.1), *Zea mays* (milho) (v6a), *Oryza sativa* (arroz) (v7) e *Brachypodium distachyon* (v3.1). Para tal, foram utilizados os métodos de busca *Blastp* e Cadeias Ocultas de Markov (*Hidden Markov Models*, HMM), sendo que o *blastp* foi realizado no banco de dados *Phytozome* 10.1 (<http://phytozome.jgi.doe.gov/pz/portal.html>) (GOODSTEIN *et al.*, 2012) e o HMM por meio do software *HMMer* v3.1 (<http://hmmer.janelia.org/>) (FINN *et al.*, 2015), ambos tendo como base de dados o proteoma disponível de cada espécie. O HMM foi realizado por meio de dois *scripts* distintos, o *hmmesbuild* que constrói um perfil com base nas proteínas fornecidas, e o *hmmesearch* que realiza a busca por probabilidade a partir deste perfil em um banco de dados de interesse. Adicionalmente, foi utilizado o *script jackhmmes* a fim de se realizar uma busca complementar, entretanto essa busca não retornou novos resultados. Após essa cobertura metodológica, as sequências obtidas e não redundantes foram reunidas para posterior análise.

Todas as sequências proteicas identificadas foram anotadas para a presença dos domínios característicos de suas respectivas enzimas membro através do *PFAM* (FINN *et al.*, 2016) (<http://pfam.sanger.ac.uk>). Apenas as sequências com identificação positiva dos domínios característicos das enzimas AMP_binding (4CL), p450 (C3H, C4H e F5H), ADH_N (CAD), Epimerase (CCR), Methyltransf_3 (CCoAOMT), Methyltransf_2 (COMT), Transferase (HCT) e Lyase_aromatic (PAL) foram anotadas posteriormente para as seguintes características: código da sequência; sequência primária; nome (ou alias, quando houver); localização cromossomal; sentido da sequência (adotou-se sinal positivo (+) para a orientação 5'→3' e (-) para o inverso); tamanho da sequência proteica; se a

sequência proteica está completa ou incompleta e; ainda, domínio da proteína e seu *e-value*. As sequências proteicas e codificantes foram baixadas e armazenadas para posteriores análises.

Para cana-de-açúcar (*Saccharum* spp.), devido à ausência de genoma e de uma plataforma de fácil uso como o *Phytozome* para a busca de suas sequências codificantes, proteicas e informações relevantes, o processo de identificação e anotação seguiu a seguinte plataforma. Bibliotecas de RNA-Seq de raiz (informação pessoal¹), folha (CARDOSO-SILVA *et al.*, 2014), entrenó (VICENTINI *et al.*, 2015), plântula (informação pessoal²) e SUCEST (VETTORE *et al.*, 2003) foram concedidas pelos colaboradores Dr. Renato Vicentini (UNICAMP), Dr. Marcos Buckeridge (USP) e Dra. Adriana Hemerly (UFRJ) para a identificação e anotação das sequências homólogas das famílias gênicas e das enzimas envolvidas na via de biossíntese de lignina. Para tal, as sequências codificantes das espécies *Sorghum bicolor* (sorgo) e *Zea mays* (milho) foram usadas como isca para a busca das sequências homólogas em cana-de-açúcar por meio do método *blastn*. Todas as sequências foram submetidas à análise da presença dos domínios característicos de cada família gênica ao *PFAM*. Todas as sequências positivas foram traduzidas em aminoácidos através do software *ExPASy translate* (<http://web.expasy.org/translate/>) (GASTEIGER *et al.*, 2003) para a identificação correta da fase de leitura aberta (*ORF* – *Open Read Frame*) novamente validada para a presença do domínio característico no *PFAM*. Após estas confirmações, as sequências proteicas positivas foram reunidas em um banco de dados para posterior análise.

3.2 Alinhamentos e Análises filogenéticas

Todas as sequências proteicas de *Arabidopsis*, sorgo, milho, arroz e *brachypodium* dentro de suas respectivas famílias gênicas foram submetidas ao alinhamento múltiplo por meio do software *MUSCLE* (<http://www.ebi.ac.uk/Tools/msa/muscle/>) (Edgar, 2004). As árvores filogenéticas foram realizadas através do MEGA 6.0 (Molecular Evolutionary Genetics Analysis) (TAMURA *et al.*, 2013) com os parâmetros Bootstrap, 1000 replicações, método de *p-distance* e deleção par a par.

Para cana-de-açúcar, em virtude da presença de sequências incompletas e de tamanho reduzido, o que inviabiliza a utilização do software MEGA, foi utilizado o

¹ Dr. Marcos Buckeridge (USP/SP), 2015.

² Dra. Adriana Hemerly (UFRJ), 2015.

software *Seaview* v.4.5.4 (GOUY; GUINDON; GASCUEL, 2010). A construção das árvores filogenéticas foi feita pelo *PhyML* do *Seaview* utilizando o método de análise por máxima verossimilhança para as sequências proteicas de cana-de-açúcar, sorgo, milho, arroz e brachypodium. Para a construção da filogenia pelo método de máxima verossimilhança, realizou-se o *Model Test* para a predição do melhor modelo estatístico.

As sequências obtidas foram alinhadas com os prováveis ortólogos de sorgo por meio do software *BioEdit Sequence Alignment Editor* v.7.2.5 para se verificar se estavam completas ou incompletas. Aquelas proteínas que mais se assemelham às *bona fide* de sorgo e que estão mais completas em relação a estas foram alinhadas na função *blast2seq* do *National Center for Biotechnology Information* (NCBI) para obtenção de informações de identidade e cobertura.

4 RESULTADOS E DISCUSSÃO

4.1 Identificação e anotação dos genes de *Arabidopsis* e gramíneas

A partir do método de busca blastp e HMM, utilizaram-se os genes já identificados e anotados em *Arabidopsis* (RAES *et al.*, 2003) como isca e foram encontrados 448 genes com domínio característico às oito famílias gênicas que incluem as enzimas da via de biossíntese de lignina. Desses, apenas 63 já haviam sido previamente identificados (nove *CADs*, sete *CCoAOMTs*, treze *4CLs*, sete *CCRs*, quatro *PALs*, uma *C4H*, uma *HCT*, dezesseis *COMTs*, três *C3Hs* e duas *F5Hs*) (RAES *et al.*, 2003; XU *et al.*, 2009), revelando uma grande quantidade de genes candidatos de estudos até então desconhecidos. Após a busca pelo método blast e HMM, foram encontrados 640, 598, 874 e 563 genes em sorgo, milho, arroz e *brachypodium*, respectivamente.

Em *Arabidopsis*, quatro genes foram encontrados na família Lyase_aromatic, correspondendo àqueles já descritos em literatura (RAES *et al.*, 2003) como codificantes de enzimas PAL. Isso não chega a ser surpreendente, visto que em todas as espécies já estudadas a família PAL é representada por um pequeno número de genes (OLSEN *et al.*, 2008). Sendo assim, todos os possíveis candidatos já haviam sido identificados. Foram encontrados oito em sorgo, quatorze em milho, treze em arroz e oito em *brachypodium*. Para todas as espécies exceto arroz, essa é a família com menor quantidade de genes.

Na família p450, a mais numerosa, 216 genes foram encontrados em *Arabidopsis*, sendo que seis deles já haviam sido identificados. Essa família, além de ser mais numerosa *per se*, agrupa três enzimas da via, devido ao seu caráter semelhante. Tanto a C3H quanto a F5H e a C4H apresentam o mesmo domínio (p450), de maneira que, quando se realiza a busca de cada família separada, os resultados se sobrepõem quase que inteiramente. Devido a esse domínio em comum e à sobreposição de resultados, foram agrupadas e tratadas como uma só família nessa etapa do trabalho. Em sorgo foram encontrados 342 genes, em milho 274, em arroz 401 e em *brachypodium* 269. Para todas as espécies, essa é a família mais numerosa.

Na família AMP-binding foram encontrados quarenta e cinco representantes em *Arabidopsis*, dos quais 13 já haviam sido identificados (RAES *et al.*, 2003); 49 em sorgo, 57 em milho, 78 em arroz e 46 em *brachypodium*. Uma das famílias com menor número de

representantes, a Methyltransf_3 apresentou sete resultados na pesquisa em Arabidopsis, e oito, seis, quatorze e dez em sorgo, milho, arroz e brachypodium, respectivamente.

Para a ADH_N, 40 genes foram anotados em Arabidopsis, nove deles já descritos previamente, 32 em sorgo, 39 em milho, 68 em arroz e 31 em brachypodium. Para a família Methyltransf_2, 17 genes foram encontrados em Arabidopsis, sendo 14 já anotados, e 32, 39, 68 e 31 em sorgo, milho, arroz e brachypodium, respectivamente.

Em Arabidopsis, 63 genes foram encontrados na família Transferase, sendo um já descrito, 94 em sorgo, 100 em milho, 140 em arroz e 95 em brachypodium. Para a família Epimerase, foram encontrados 56 em Arabidopsis, com sete descritos; 60 em sorgo, 80 em milho, 115 em arroz e 80 em brachypodium.

Esses dados estão sumarizados na tabela 3. A quantidade de sequências não apresenta relação direta com a quantidade de enzimas envolvidas na via de biossíntese de lignina, pois esses domínios são envolvidos em diversos processos naturais do organismo. O maior exemplo é o domínio p450, presente nas enzimas C3H, C4H e F5H, porém envolvido também em diversas funções diferentes nas plantas (LIU *et al.*, 2015), enquanto que os integrantes propriamente funcionais da via são geralmente um ou dois por enzima, variando conforme a espécie. Já para o domínio Lyase_aromatic, presente na enzima PAL, grande parte das sequências que possuem esse domínio está relacionados em algum nível à síntese de lignina.

Tabela 3 - Número de genes das famílias gênicas que contém membros envolvidos na biossíntese de lignina

Enzima	Família	Espécie				
		Arabidopsis	Sorgo	Milho	Arroz	Brachypodium
4CL	AMP-binding	45	49	57	78	46
C3H	p450	216	342	274	401	269
CAD	ADH_N	40	32	39	68	31
CCoAOMT	Methyltransf_3	7	8	6	14	10
CCR	Epimerase	56	60	80	115	80
COMT	Methyltransf_2	17	47	28	45	24
HCT	Transferase	63	94	100	140	95
PAL	Lyase_aromatic	4	8	14	13	8

Fonte: do autor.

4.2 Identificação e anotação das sequências de cana-de-açúcar

Na busca de sequências de cana-de-açúcar foram utilizados cinco bancos de transcriptoma (Tabela 4), tais como o SUCEST e quatro bibliotecas de RNA-Seq desenvolvidas a partir de raiz, folha, entrenó e plântula. A quantidade de sequências obtidas nos diferentes transcriptomas apresenta um padrão similar e próximo a quantidade de genes nas espécies com genomas disponível estudados (Tabela 3). Por outro lado, uma análise comparativa dentre os transcriptomas, observou-se grande variação numérica. O banco de raiz resultou num total de 205 sequências, enquanto que o banco de plântula resultou em 603. Alguns dos bancos têm como característica a existência de sequências mais curtas, ou seja, apenas fragmentos de sequências gênicas, o que pode impactar no total de sequências não redundantes.

Tabela 4 - Número de genes das famílias gênicas relacionadas com a biossíntese de lignina nas diferentes bibliotecas de cana-de-açúcar.

Enzima	Família	Bibliotecas				
		Raiz	Folha	Entrenó	Plântula	SUCEST
4CL	AMP-binding	36	62	55	72	58
C3H	p450	75	223	148	289	90
CAD	ADH_N	34	38	37	53	51
CCoAOMT	Methyltransf_3	6	9	6	8	12
CCR	Epimerase	14	72	30	54	36
COMT	Methyltransf_2	12	22	16	39	15
HCT	Transferase	23	78	56	80	73
PAL	Lyase_aromatic	5	8	6	8	9
Total		205	512	354	603	344

Fonte: do autor.

Dentre essas bibliotecas, apenas duas possuem estudos publicados acerca dos genes envolvidos na biossíntese de lignina. No trabalho realizado por Bottcher e colaboradores (2013), a partir do banco SUCEST foram descritas 33 sequências relacionadas à via de biossíntese de lignina. Dessas, todas foram encontradas no presente trabalho. No trabalho de Vicentini e colaboradores (2015), foram identificados 21 transcritos que codificam proteínas envolvidas na biossíntese de lignina.

4.3 Identificação dos clados *bona fide* da via de biossíntese de lignina em gramíneas

Tendo em mãos o conjunto de proteínas nas espécies com genoma sequenciado, realizou-se a filogenia através do método *Neighbor-joining* para cada família gênica, a fim de avaliar a relação filogenética entre as espécies e encontrar os genes chamados *bona fide*, ou seja, aqueles que podem de fato estar envolvidos na via de biossíntese de lignina (Figuras 6-13). Para a família gênica Epimerase, foram retiradas sequências pertencentes ao clado A da classificação de Yin e colaboradores (2011), visto que as enzimas pertencentes a este clado apresentam funções relacionadas à biossíntese de hemicelulose, grande quantidade de sequências dificultando a visualização e identificação dos códigos gênicos na árvore filogenética. Assim, a figura 10 representa as epimerases pertencentes aos clados B e C, de acordo com Yin e colaboradores (2011).

A seleção do clado *bona fide* se baseou na proximidade filogenética com as sequências de *Arabidopsis* e nos dados encontrados na literatura acerca dessas proteínas. Recentemente, Carocha e colaboradores (2015) buscando a identificação dos genes da via de biossíntese de lignina em eucalipto, realizaram a identificação por homologia destes genes em gramíneas como arroz e milho, sendo que para o milho os autores estudaram apenas os genes *COMT* (*ZmaCOMT*) e *CCR* (*ZmaCCR1* e *ZmaCCR2*). Das sequências encontradas por esse para estas duas espécies, todas foram inclusas nos respectivos clados *bona fide*, o que consolida a robustez do método adotado.

Figura 6 - Árvore filogenética das proteínas membro da família gênica AMP_binding e do clado 4CL nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica AMP e B) Filogenia do clado das proteínas 4CL. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

Figura 7 - Árvore filogenética das proteínas membro da família gênica p450 e dos clados C3H, C4H e F5H nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica p450; B) Filogenia do clado C3H; C) Filogenia do clado C4H e; D) Filogenia do clado F5H. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". As proteínas que contêm o domínio p450 que não se enquadram nos clados de interesse foram colapsadas a fim de permitir uma melhor visualização.

Fonte: do autor.

Figura 8 - Árvore filogenética das proteínas membro da família gênica ADH_N e do clado CAD nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica ADH_N e B) Filogenia do clado CAD. Os códigos dos genes extraídos do Phytosome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

Figura 9 - Árvore filogenética das proteínas membro da família gênica Methyltransf_3 e do clado CCoAOMT nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica Methyltransf_3 e B) Filogenia do clado CCoAOMT. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

Figura 10 - Árvore filogenética das proteínas membro da família gênica Epimerase (clados B e C) e do clado CCR nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica Epimerase e B) Filogenia do clado CCR. Os códigos dos genes extraídos do Phytosome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

Figura 11 - Árvore filogenética das proteínas membro da família gênica Methyltransf_2 e do clado COMT nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica Methyltransf_2 e B) Filogenia do clado COMT. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

Figura 12 - Árvore filogenética das proteínas membro da família gênica Transferase e do clado HCT nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

A) Filogenia da família gênica Transferase e B) Filogenia do clado HCT. Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

Figura 13 - Árvore filogenética das proteínas membro da família gênica Lyase_aromatic, enzimas PAL, nas espécies *A. thaliana*, *B. distachyon*, *O. sativa*, *S. bicolor* e *Z. mays*.

Os códigos dos genes extraídos do Phytozome para as diferentes espécies sofreram alteração para se adaptarem a figura. Para *Brachypodium*, o código "Bradi" foi modificado para "Bd"; para *Oryza sativa* o código "LOC_Os" foi modificado para "Os"; para *Z. mays* o código "GRMZM" foi modificado para "Zm" e para *S. bicolor* o código foi alterado de "Sobic." para "Sb". Fonte: do autor.

A partir das filogenias entre as gramíneas e *Arabidopsis* foi possível selecionar o clado *bona fide* envolvido na via de biossíntese de lignina de cada família gênica. Estes clados foram selecionados baseado das sequências identificadas na literatura como envolvidas na síntese de lignina nas gramíneas e *Arabidopsis*. Em virtude da topologia do arranjo dos ramos na filogenia da família Lyase_aromatic (enzima PAL), não foi possível identificar um clado entre *Arabidopsis* e gramíneas com ortologia estrutural.

A enzima CAD tem sido amplamente estudada por seu papel na lignificação. O estudo com mutantes pode auxiliar na descoberta de genes importantes e as repercussões das alterações nesses genes para a planta. Plantas transgênicas de milho com atividade reduzida de ZmCAD2 (Zm5G844562) por meio de RNA de interferência apresentam um aumento na disponibilidade de polissacarídeos da parede (20% a mais de celulose e 30% de xilose) e um subsequente aumento na degradabilidade e na produção de bioetanol sem efeitos negativos no crescimento e desenvolvimento da planta (FORNALE *et al.*, 2012). Dentre 12 genes CAD de arroz, o silenciamento do gene *bona fide OsCAD2* (Os02g09490) ortólogo a *ZmCAD2* resultou na redução pouco expressiva no conteúdo de lignina (5 a

6%), o que indica a possibilidade da importância de genes que não estão localizados no clado *bona fide* (HIRANO *et al.*, 2012). Dados de cinética da conversão de substratos a produtos mostram que a proteína de brachypodium BdCAD5 (Bradi3g06480, ortólogo a *ZmCAD2* e *OsCAD2*), expresso em todos os tecidos e num padrão semelhante a *OsCAD2*, é a principal CAD na síntese de lignina em brachypodium (BUKH *et al.*, 2012; TOBIAS e CHOW, 2005), suportando a função CAD para os genes membros do clado *bona fide*.

Adicionalmente, o mutante *bml* milho sem alteração da anatomia da planta e com o conteúdo total de lignina reduzido a 20% foi geneticamente mapeada para o gene *ZmCAD2* (HALPIN *et al.*, 1998; YOON *et al.*, 2015). O estudo do mutante de sorgo *Bmr6*, que altera a expressão da *SbCAD4* (Sb004G071000, ortólogo a *ZmCAD2*), mostrou que essa CAD é a principal em sorgo e apresenta 97,8% de similaridade com a *ZmCAD2*. O conteúdo de lignina é significativamente reduzido no mutante nulo desse gene (SATTTLER *et al.*, 2009; SATTTLER *et al.*, 2010). O silenciamento de *PviCAD2* (Pavir.J34526) em *switchgrass* causa diminuição da lignificação e aumento na sacarificação (SAATHOFF *et al.*, 2011). Todos esses alelos identificados representam ferramentas para manipulação da via de biossíntese de lignina.

Em um estudo sobre resposta a estresse hídrico foi relatado o padrão de expressão similar entre os genes de arroz *OsCAD2* e *OsCOMT* (Os08g06100) (HU *et al.*, 2009), o que sugere um sincronismo desses genes durante a biossíntese de lignina. O gene *SbCOMT* (Sb007G047300) é tido como candidato ideal para manipulação da síntese de monolignóis, baseado nos efeitos observados nos mutantes *bmr12* como a diminuição da relação S/G e alteração no conteúdo de lignina (GREEN *et al.*, 2014). O gene *bmr12* de sorgo codifica a enzima *SbCOMT* (Sb007G047300), assim como o *Bm3* de milho codifica *ZmCOMT* (Zm196475, ortólogo de *SbCOMT*) (SATTTLER *et al.*, 2014).

Dentre as quatro HCTs de arroz (OsHCT1-4), a proteína OsHCT1 (Os04g42250) é a mais expressa em todos os tecidos (KIM *et al.*, 2012), indicando que esta enzima é a principal HCT desta espécie.

O mutante *bmr2* de sorgo codifica a enzima 4CL, e o fenótipo desse mutante evidencia o papel desta 4CL na lignificação (SABALLOS *et al.*, 2012). O mutante com expressão reduzida no gene *Os4CL3* (Os02g08100) de arroz apresenta crescimento reduzido, apesar de não afetar outras características agrônômicas como tempo de floração e comprimento da folha (GUI *et al.*, 2011). Em milho, comparando-se diferentes estágios de

desenvolvimento por meio da análise de expressão nos entrenós 9 e 13 (IN9 e IN13, respectivamente), foi observado que duas 4CLs apresentaram maior expressão no IN9 (Zm2G019746, Zm2G075333) e duas foram mais expressas no IN13 (Zm2G054013, Zm2G080663) (BOSCH *et al.*, 2011).

O mesmo trabalho (BOSCH 2011) mostra que três CCoAOMTs (Zm2G033952 - *ZmCCoAOMT1*, Zm2G004138 - *ZmCCoAOMT-like1* e Zm2G099363 - *ZmCCoAOMT2*) apresentaram maior expressão no IN9, bem como uma CCR (Zm2G131205). O silenciamento de *CCR1* de milho aumenta a digestibilidade, sem comprometer o crescimento e desenvolvimento da planta (TAMASLOUKHT *et al.* 2011). Quatro enzimas PAL exibiram uma expressão diferenciada no IN9 (Zm2G160541, Zm2G074604, Zm2G029048, Zm2G081582), sendo a expressão PAL em arroz relacionada à resposta a estresse (DUAN *et al.*, 2014), sugerindo um papel da lignina no mecanismo de resistência presente em variedades naturalmente menos suscetíveis a estresse.

De forma geral, baseado nas informações disponíveis para *Arabidopsis*, *brachypodium*, sorgo, milho e arroz, foi possível identificar as sequências *bona fide* envolvidas na biossíntese de lignina de cada família gênica nas análises filogenéticas entre as sequências proteicas dessas espécies. Essa abordagem permitiu identificar os clados chamados *bona fide* contendo as enzimas envolvidas e possivelmente envolvidas na biossíntese de lignina (Figuras 6 a 13).

4.4 Identificação das sequências de cana-de-açúcar pertencentes aos clados *bona fide* e definição dos genes candidatos

O grupo de ortólogos de milho e sorgo pertencentes aos clados *bona fide* envolvidos na biossíntese de lignina foram usados para a construção de filogenias através do método de máxima verossimilhança com todas as sequências de cana-de-açúcar identificadas nas cinco bibliotecas transcriptômicas dentro de cada família gênica por meio do software *Seaview* (dados não mostrados). Esta abordagem permitiu identificar as sequências de cana-de-açúcar ortólogas as sequências de sorgo e milho possivelmente envolvidas na biossíntese de lignina.

As sequências encontradas no banco SUCEST que se localizaram nesses clados *bona fide* somam 56 ao total, sendo duas *4CLs*, três *CADs*, quatro *COMTs*, oito *PALs*, onze *CCoAOMTs*, onze *CCRs*, cinco *HCTs*, duas *C3Hs*, sete *C4Hs* e três *F5Hs*. É interessante notar que, das 33 previamente descritas no SUCEST (Bottcher *et al.*, 2013), 22 foram localizadas nos clados *bona fide*, sendo uma *4CL*, duas *CADs*, duas *COMTs*, quatro *PALs*, seis *CCoAOMTs*, duas *CCRs*, uma *HCT*, duas *C3Hs*, uma *C4H* e uma *F5H*. Dessa forma, foram encontradas 34 novas sequências até então não identificadas anteriormente, sendo elas uma *4CL*, uma *CAD*, duas *COMTs*, quatro *PALs*, cinco *CCoAOMTs*, nove *CCRs*, quatro *HCTs*, seis *C4Hs* e duas *F5Hs*. Exceto para a enzima *C3H*, todas as demais apresentam resultados que até então não haviam sido identificados, em especial *CCR*, *C4H*, *CCoAOMT* e *PAL*, com elevado número de novas sequências. Esta grande quantidade de novas sequências permite expandir a lista de resultados obtidos desse banco e expõe a vantagem da abordagem explorado no presente trabalho.

O conjunto de sequências de cana-de-açúcar que divergiu a partir do ancestral comum dentre as gramíneas membro do clado *bona fide* foram selecionadas e alinhadas com a sequência ortóloga de sorgo. Esses alinhamentos são úteis para visualização do grau de conservação, além de informarem quão completas estão as sequências em relação a seu ortólogo em sorgo (Figuras 14-23).

Dentre os sete genes de sorgo membros do clado *4CL* (Figura 6), cinco genes (Sb007G145600, Sb010G214900, Sb004G062500, Sb007G089900 e Sb004G272700) tiveram grupos de sequências ortólogas (COGs – *cluster of orthologous genes*) em cana-de-açúcar. No alinhamento de cada COG *4CL* (figura 14), foi possível encontrar sequências completas (C), incompletas (A, B e D) e ainda sequências de diferentes bancos que, se sobrepostas, são capazes de se equivar a todo o comprimento da sequência de sorgo (E).

Figura 14 - Alinhamento de sequências proteicas de 4CL dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

(Continua)

Figura 14 - Continuação

B

B

```

                                     10      20      30      40      50      60      70      80
Sb010G214900  MGSVAEDSAAPAAASVVFRSKLPDIEIPRHLSLQAYCFERLPEVSSRPCLIDGQTGAVHTYADVERLSRTAAAAALRGLGVGKGDVVMN
Sc_4CL_2      -----
Rt_4CL_1      -----
Sc_4CL_3      -----
En_4CL_1      -----
Rt_4CL_2      -----
Fl_4CL_2      -----

                                     110      120      130      140      150      160      170      180
Sb010G214900  LGAARLGAATTTANPFYTPHEIRQAAAAAGAKVIVTEACAVEKVRGFAAERGVPV~ATVDGGAFDGCLELGALMDAAEPLADDEEVDP
Sc_4CL_2      -----AGAKVIVTEACAVEKVRGFAAERGVPVAV~DG~AFEGCLELGALMDAEPLADDEEVDP
Rt_4CL_1      -----AAEPLADDEEVDP
Sc_4CL_3      -----
En_4CL_1      -----
Rt_4CL_2      -----
Fl_4CL_2      -----

                                     210      220      230      240      250      260      270      280
Sb010G214900  TGMPKGVMLTHRSLVTSVAQQVDGENPNLHFSSDDVVLCVLPLPHITYSLNSVLLAGLRAGCAIVIMRKFEIGALVELVRAHGVTVAPF
Sc_4CL_2      TGMPKGVMLTHRSLVTSVAQQVDGENPNLYFSSDDVVLCVLPLPHITYSLNSVLLAGLRAGCAIVIMRKFEIGALVELVRAHGVTVAPF
Rt_4CL_1      TGMPKGVMLTHRSLVTSVAQQVDGENPNLYFSSDDVVLCVLPLPHITYSLNSVLLAGLR
Sc_4CL_3      -----WSVCAIFQVDGENPNLYFSSDDVVLCVLPLPHITYSLNSVLLAGLRAGCAIVIMRKFEIGALVELVRAHGVTVAPF
En_4CL_1      -----APF
Rt_4CL_2      -----
Fl_4CL_2      -----

                                     310      320      330      340      350      360      370      380
Sb010G214900  RVGAADLASIRMVMSGAAPMGKDLQDAFMAKIPNAVLGQGYGMTEAGFVLAMCLAFAKEPFEVKSGSCGTVVRNAELKIVDPDTSESL
Sc_4CL_2      RVGAHDLASIRMVMSGAAPMGKDLQDAFMAKIPNAVLGQGYGMTEAGFVLAMCLAFAKEPFEVKSGSCGTVVRNAELKIVDPDTSASL
Rt_4CL_1      -----
Sc_4CL_3      RVGAHDLASIRMVMSGAAPMGKDLQDAFMAKIPNAVLGQXYGMTEAGFVLAMCLAFAKEPFEVKSRFLRRHQERGA
En_4CL_1      RVTAADLASIRMVMSGAAPMGKDLQDAFMAKIPNAVLGQGYGMTEAGFVLAMCLAFAKEPFEVKSGSCGTVVRNAELKIVDPDTSASL
Rt_4CL_2      -----MVMSGAAPMGKDLQDAFMAKIPNAVLGQGYGMTEAGFVLAMCLAFAKEPFEVKSGSCGTVVRNAELKIVDPDTSASL
Fl_4CL_2      -----MSGAAPMGKDLQDAFMAKIPNAVLGQGYGMTEAGFVLAMCLAFAKEPFEVKSGSCGTVVRNAELKIVDPDTSASL

                                     410      420      430      440      450      460      470      480
Sb010G214900  BQIMKGYLNDPEATKNTIDKGWLHTGDIGYVDDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEIKDAAVVSMKDELAGEVEV
Sc_4CL_2      BQIMKGYLNDPEATKNTIDKGWLHTGDIGYVDDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEIKDAAVVSMKDELAGEVEV
Rt_4CL_1      -----
Sc_4CL_3      -----
En_4CL_1      BQIMKGYLNDPEATKNTIDKGWLHTGDIGYVDDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEIKDAAVVSMKDELAGEVEV
Rt_4CL_2      BQIMKGYLNDPEATKNTIDKGWLHTGDIGYVDDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEIKDAAVVSMKDELAGEVEV
Fl_4CL_2      BQIMKGYLNDPEATKNTIDKGWLHTGDIGYVDDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEIKDAAVVSMKDELAGEVEV

                                     510      520      530      540      550
Sb010G214900  ENEIKQFVAKEVVFYKRINRVFFTDSIPKNPSGKILRKDLRARLAAGIPSSDNTQSKS
Sc_4CL_2      ENEIKQFVAKEVVFYKRINRVFFTDSIPKNPSGKILRKDLRARLAAGIPSSDNTQSKS
Rt_4CL_1      -----
Sc_4CL_3      -----
En_4CL_1      ENEIKQFVAKEV
Rt_4CL_2      ENEIKQFVAKEVVFYKRINRVFFTDSIPKNPSGKILRKDLRARLAAGIPSSDNTQSKS
Fl_4CL_2      ENEIKQFVAKEVVFYKRINRVFFTDSIPKNPSGKILRKDLRARLAAGIPSSDNTQSKS

```

(Continua)

(Continua)

Figura 14 - Continuação

C

```

Sb004G062500 10 20 30 40 50 60 70 80
Rt_4CL_3 MGSVDTA VAVPVPVPEPAEEKAAVFRSKLPDIEINNSQSLQTYCFGKMS EVADRA CLIDGQTG ASYTYAEVESLSRRAASGLRAMG
P1_4CL_2 MGSVDTA VAVPVPVPEPAEEKAAVFRSKLPDIEINNSQSLHAYCFGKMS EVADRA CLIDGQTG ASYTYAEVESLSRRAASGLRAMG
Sc_4CL_4 MGSVDTA VAVPVPVPEPAEEKAAVFRSKLPDIEINNSQSLHAYCFGKMS EVADRA CLIDGQTG ASYTYAEVESLSRRAASGLRAMG
Sc_4CL_5 MGSVDTA VAVPVPVPEPAEEKAAVFRSKLPDIEINNSQSLHAYCFGKMS EVADRA CLIDGQTG ASYTYAEVESLSRRAASGLRAMG
En_4CL_2 MGSVDTA VAVPVPVPEPAEEKAAVFRSKLPDIEINNSQSLHAYCFGKMS EVADRA CLIDGQTG ASYTYAEVESLSRRAASGLRAMG
F1_4CL_3 MGSVDTA VAVPVPVPEPAEEKAAVFRSKLPDIEINNSQSLHAYCFGKMS EVADRA CLIDGQTG ASYTYAEVESLSRRAASGLRAMG

110 120 130 140 150 160 170 180
Sb004G062500 NCPEFAFTFLGAARLGAATTTANPFYTPHEIHRQAEAAAGAKVIVTEACAVEKRVREFAAGRGVFVVTV DGRFDGCVFAEVI--AAEELE
Rt_4CL_3 NCPEFAFTFLGAARLGAATTTANPFYTPHEIHRQAEAAAGAKLIVTEACAVEKRVREFAAGRGVFVVTV DGRFDGCVFAEVI--AAEELE
P1_4CL_2 NCPEFAFVFLGAARLGAATTTANPFYTPHEIHRQAEAAAGAKLIVTEACAVEKRVREFAAGRGVFVVTV DGRFDGCVFAEVI--AAEELE
Sc_4CL_4 NCPEFAFTFLGAARLGAATTTANPFYTPHEIHRQAEAAAGAKLIVTEACAVEKRVREFAAGRGVFVVTV DGRFDGCVFAEVI--AAEELE
Sc_4CL_5 NCPEFAFTFFGAARLGAATTTGNPFYTPHEIH-----
En_4CL_2 NCPEFAFTFLGAARLGAATTTANPFYTPHEIHRQAEAAAGAKLIVTEACAVEKRVGFAAERGVFVVAVDGA FEGCLEL GALMDAAEPLA
F1_4CL_3 NCPEFAFTFLGAARLGAATTTANPFYTPHEIHRQAEAAAGAKLIVTEACAV-----

210 220 230 240 250 260 270 280
Sb004G062500 LPYSSGTTGLPKGVMLTHRS LITSVAQQVDGENPNLYFSKDDVLLCLLP LPHIYSLNSVLLAGLRAGSTIVIMRKFDL GALV DLVRKH
Rt_4CL_3 LPYSSGTTGLPKGVMLTHRS LITSVAQQVDGENPNLYFSKDDVLLCLLP LPHIYSLNSVLLAGLRAGSTIVIMRKFDL GALV DLVRKH
P1_4CL_2 LPYSSGTTGLPKGVMLTHRS LITSVAQQVDGENPNLYFSKDDVLLCLLP LPHIYSLNSVLLAGLRAGSTIVIMRKFDL GALV DLVRKH
Sc_4CL_4 LPYSSGTTGLPKGVMLTHRS LITSVAQQVDGENPNLYFSKDDVLLCLLP LPHIYSLNSVLLAGLRAGSTIVIMRKFDL GALV DLVRKH
Sc_4CL_5 LPYSSGTTGLPKGVMLTHRS LITSVAQQVDGENPNLYFSKDDVLLCLLP LPHIYSLNSVLLAGLRAGSTIVIMRKFDL GALV DLVRKH
En_4CL_2 LPYSSGTTGLPKGVMLTHRS LITSVAQQVDGENPNLYFSKDDVLLCLLP LPHIYSLNSVLLAGLRAGSTIVIMRKFDL GALV DLVRKH
F1_4CL_3 -----

310 320 330 340 350 360 370 380
Sb004G062500 VEIAKSPRVTAADLASIRMVMSGAAPMGKELQDAFMTKIPNAVLGQGYG MTEAGFVLAMCLAFAPKPFQVKSGSGCTVVRNAELKIVD
Rt_4CL_3 VEIAKSPRVTAADLASIRMVMSGAAPMGKELQDAFMTKIPNAVLGQGYG MTEAGFVLAMCLAFAPKPFQVKSGSGCTVVRNAELKIVD
P1_4CL_2 VEIAKSPRVTAADLASIRMVMSGAAPMGKELQDAFMTKIPNAVLGQGYG MTEAGFVLAMCLAFAPKPFQVKSGSGCTVVRNAELKIVD
Sc_4CL_4 VEIAKSPRVTAADLASIRMVMSGAAPMGKELQDAFMTKIPNAVLGQGYG MTEAGFVLAMCLAFAPKPFQVKSGSGCTVVRNAELKIVD
Sc_4CL_5 VEIAKSPRVTAADLASIRMVMSGA-----
En_4CL_2 VEIAKSPRVTAADLASIRMVMSGA-----
F1_4CL_3 -----

410 420 430 440 450 460 470 480
Sb004G062500 GEICIRGEQIMKGYLNDLESTKNTIDKGWLHTGDIGYVDDDDIEFIVDR LKEIIRYKGFQVPPAELEALLITHPEIKDAAVVSMKDD
Rt_4CL_3 GEICIRGEQIMKGYLNDLESTKNTIDKGWLHTGDIGYVDDDDIEFIVDR LKEIIRYKGFQVPPAELEALLITHPEIKDAAVVSMKDD
P1_4CL_2 GEICIRGEQIMKGYLNDLESTKNTIDKGWLHTGDIGYVDDDDIEFIVDR LKEIIRYKGFQVPPAELEALLITHPEIKDAAVVSMKDD
Sc_4CL_4 GEICIRGEQIMKGYLNDLESTKNTIDKGWLHTGDIGYVDDDDIEFIVDR LKEIIRYKGFQVPPAELEALLITHPEIKDAAVVSMKDD
Sc_4CL_5 GEICIRGEQIMKGYLNDLESTKNTIDKGWLHTGDIGYVDDDDIEFIVDR LKEIIRYKGFQVPPAELEALLITHPEIKDAAVVSMKDD
En_4CL_2 -----
F1_4CL_3 -----

510 520 530 540 550
Sb004G062500 TEGSEVTEDEIKQFVAKEVVFYKKVHKVFFTESIPKNPSGKI LRKDLRARLAAGVH
Rt_4CL_3 TEGSEVTEDEIKQFVAKEVVFYKKVHKVFFTESIPKNPSGKI LRKDLRARLAAGVH
P1_4CL_2 TEGSEVTEDEIKQFVAKEVVFYKKVHKVFFTESIPKNPSGKI LRKDLRARLAAGVH
Sc_4CL_4 TEGSEVTEDEIKQFVAKEVVFYKKVHKVFFTESIPKNPSGKI LRKDLRARLAAGVH
Sc_4CL_5 TEGSEVTEDEIKQFVAKEVVFYKKVHKVFFTESIPKNPSGKI LRKDLRARLAAGVH
En_4CL_2 -----
F1_4CL_3 -----

```

(Continua)

(Continua)

Figura 14 - Continuação

D

	10	20	30	40	50	60	70	80	
Sb007G089900	MGS TE EEQQQPVFVVA AAV EEEEASPEIIFRSKLPDIAITNPLPLHRYCFERLPEVADRPC LID GATGAVRTYA EV DLTRRLAAALRR								
En_4CL_3	-----								
P1_4CL_3	-----								
P1_4CL_4	-----								
F1_4CL_4	-----								

	110	120	130	140	150	160	170	180	
Sb007G089900	NLMLNSAEFVLSFFAASRVGA AV TTANPMSTPHEIANQIAASGATVVFTESMAVDKLP AV QDDGGGLTVVLID AHR DGCLHFWDDVMA								
En_4CL_3	-----								
P1_4CL_3	-----								
P1_4CL_4	-----								
F1_4CL_4	-----								

	210	220	230	240	250	260	270	280	
Sb007G089900	AAAAAGFD DD VVALPYSSGTTGLPKGVMLTHRSLS TS VAAQV Q DGNPNIGFTADDVILCSLPMFHIYS LN TMCGLRVGAATVVMR								
En_4CL_3	-----								
P1_4CL_3	-----								
P1_4CL_4	-----								
F1_4CL_4	-----								

	310	320	330	340	350	360	370	380	
Sb007G089900	RHRITIAPLVPPIVVAVAKSDEAASHDLSSVRMVL SGA APMGKDIEDAFMAKLP GA VLGGYGMTEAGPVL SM CLAF AK EPFKVKS GA								
En_4CL_3	-----								
P1_4CL_3	-----								
P1_4CL_4	-----								
F1_4CL_4	-----								

	410	420	430	440	450	460	470	480	
Sb007G089900	VD PD TGKSLARNLPGEICIRGQQIMKGYLNNPDATKNSIDAGWLHTGDVGFVDDDDDEIFIV DR LKEI IK YKGLQVAPAEIEALLITH								
En_4CL_3	VD PD TGKSLARNLPGEICIRGQQIMKGYLNNPEATKNSIDAGWLHTGDVGFVDDDDDEIFIV DR LKEI IK YKGLQVAPAEIEALLITH								
P1_4CL_3	VD PD TGKSLARNLPGEICIRGQQIMKGYLNNPEATKNSIDAGWLHTGDVGFVDDDDDEIF								
P1_4CL_4	-----								
F1_4CL_4	-----								

	510	520	530	540	550	560	570	
Sb007G089900	VEPEIGEIPVAFVAKAKGSELS EDD VKQFVAKEVIYYKKVREVIFVDKIPKAPSGKILRKELRKQLQQQ--VA							
En_4CL_3	VEPEIGEIPVAFVAKAKGSELS EDD VKQFVAKEVIYYKKVREVIFVDKIPKAPSGKILRKELRKQLQQQQQVA							
P1_4CL_3	VEPEIGEIPVAFVAKAKGSELS EDD VKQFVAKEVIYYKKVREVIFVDKIPKAPSGKILRKELRKQLQQQQQVA							
P1_4CL_4	VEPEIGEIPVAFVAKAKGSELS EDD VKQFVAKEVIYYKKVREVIFVDKIPKAPSGKILRKELRKQLQQQQQVA							
F1_4CL_4	VEPEIGEIPVAFVAKAKGSELS EDD VKQFVAKEVIYYKKVREVIFVDKIPKAPSGKILRKELRKQLQQQQQVA							

(Continua)

Figura 14 - Conclusão

A) Alinhamento com Sb007G145600; B) alinhamento com Sb010G214900; C) alinhamento com Sb004G062500; D) alinhamento com Sb007G089900; E) alinhamento com Sb004G272700. As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

Para C3H, ambas as proteínas de sorgo possuem pelo menos uma sequência proteica completa encontrada de cana-de-açúcar, sendo uma identificada no SUCEST e uma no RNA-Seq de entrenó para Sb009G181800 e uma de folha para Sb003G327800, além de sequências muito próximas do comprimento da correspondente em sorgo (figura 15).

Figura 15 - Alinhamento de sequências proteicas de C3H dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

(Continua)

Figura 15 - Conclusão

A) Alinhamento com Sb009G181800; B) alinhamento com Sb003G327800. As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; F1 – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

Para C4H, apenas no alinhamento da proteína Sb002G126600 foi possível encontrar uma sequência de cana-de-açúcar completa, presente tanto no banco do SUCEST quanto nos RNA-Seqs de folha e plântula, sendo que o início da proteína Sb003G337400 não pôde ser relacionado a nenhuma sequência encontrada (Figura 16).

Figura 16 - Alinhamento de sequências proteicas de C4H dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

Figura 16 - Continuação

B

	10	20	30	40	50	60	70	80	
Sb004G141200	MAVSAARVAVATAVSLAVHLLRSFLQAQHPALGLLLAAVFLGIAATGNAGAAANAPP	GPAAPVVF	GNWLQVGN	DLNHRFLARLSARY					
Sc_C4H_3	MAVSAARVAVATAVSLAVHLLRSFLQAQHPSLGMLLPAAVFLGIAATGNADAANAPP	GPAAPVVF	GNWLQVGN	DLNHRFLARLSARY					
P1_C4H_2								GNDLNHRFLARLSARY	
	110	120	130	140	150	160	170	180	
Sb004G141200	LVVVSDPRLATEVLHTQGVDFGSRPRNVVFDIFTANGADMVFTEYGDHWRRMRVMTLPFFTARVVQQYRDMWEAEMDAVVSDISGDG								
Sc_C4H_3	LMVVSDPRLATEVLHTQGVDFGSRPRNVVFDIFTANGADMVFTEYGDHWRRMRVMTLPFFTARVVQQYRDMWEAEMDAVVSDIFADG								
P1_C4H_2	LVVVSDPRLATEVLHTQGVDFGSRPRNVVFDIFTANGADMVFTEYGDHWRRMRVMTLPFFTARVVQQYRDMWEAEMDAVVSDIFADG								
	210	220	230	240	250	260	270	280	
Sb004G141200	LQLMLYNIMYRMFDARFDSVDDPMFVEATRFNSERSRLAQSFQDYNNGDFIPILRPFRLRGYLNKCKDLQTRRLAFFNNNYVEKRRKVM								
Sc_C4H_3									
P1_C4H_2	LQLMLYNIMYRMFDAGFDSVDDPMFVEATRFNSERSRLAQSFQDYNNGDFIPILRPFRLRSYLSKCKDLQTRRLAFFNNNYVEKRRKVM								
	310	320	330	340	350	360	370	380	
Sb004G141200	IDHILQAEKSGEITSENVYIYVENINVAAIETTLWSIEWALAEVNHFAVQRKVRDEIKAVVGDEPITESTIHRLPYLQAVIKETLR								
Sc_C4H_3									
P1_C4H_2	IDHILQAEKSGEITSENVYIYVENINVAAIETTLWSIEWALAEVNHFAVQRKVRDEIKAVVGDEPITESTIHSLPYLQAVIKETLR								
	410	420	430	440	450	460	470	480	
Sb004G141200	NLEBAKLGGYTIPKSGKVVNNAWNLANNPELWEKPEEFRRPERFLGEEKSV DATVGGKVD FRFLPFGVGRRC PGIIILALPILALIVGR								
Sc_C4H_3									
P1_C4H_2	NLEBAKLGGYTIPKSGKVVNNAWNLANNPELWEKPEEFRRPERFLGEEKSV DATVGGKVD FRFLPFGVGRRC PGIIILALPILALIVGR								
	510	520	530						
Sb004G141200	VEKLDVSEKGGQFSLHIAKHSVIAFHPVSA								
Sc_C4H_3									
P1_C4H_2	VEKLDVSEKGGQFSLHIANHSVIAFHPVSP								

(Continua)

Figura16 - Conclusão

A) Alinhamento com Sb003G337400; B) alinhamento com Sb004G141200; C) alinhamento com Sb002G126600. As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

Na enzima CAD, uma sequência no RNA-Seq de entrenó se mostrou completa em relação à de sorgo Sb004G071000, sendo as demais idênticas a essa nas regiões existentes (figura 17).

Figura 17 - Alinhamento de sequências proteicas de CAD dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

No alinhamento de CCoAOMT (figura 18) pôde-se observar uma curta região de *gap* das sequências de cana-de-açúcar em relação a sorgo ao redor da posição 20, porém é interessante notar que as três sequências analisadas (duas do banco SUCEST e uma do RNA-Seq de raiz) se mostram idênticas.

Figura 18 - Alinhamento de sequências proteicas de CCoAOMT dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

Como o alinhamento de CCR foi realizado com quatro diferentes sequências de sorgo (figura 19), vários resultados se apresentam: sequência quase completa em relação a Sb004G065600, apenas com porção C-terminal ausente; três sequencias idênticas obtidas de bancos diferentes (SUCEST e nos RNA-Seqs de plântula e folha,) com pequenos gaps em relação a Sb007G141200; uma sequência praticamente idêntica a Sb002G146000 (98% de similaridade) e; sequências curtas que sobrepõem ao longo da sequência de sorgo Sb010G066000.

Figura 19 - Alinhamento de sequências proteicas de CCR dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

Figura 20 - Alinhamento de sequências proteicas de COMT dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados:

Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

As sequências encontradas em cana-de-açúcar e alinhadas às F5H de sorgo apresentaram cerca de 230 aminoácidos ausentes na região C-terminal (Figura 21). Para ambas as sequências correspondentes em sorgo (Sb001G196300 e Sb005G088400), não foi possível encontrar todas as regiões a partir das sequências de cana-de-açúcar.

Figura 21 - Alinhamento de sequências proteicas de F5H dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

A) Alinhamento com Sb001G196300; B) alinhamento com Sb005G088400. As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

Em HCT, novamente foram encontradas três sequências idênticas de bancos diferentes, que apresentam apenas um *gap* na posição 260 em relação à de sorgo, Sb004G212300. Para Sb006G136800, a região N-terminal com 74 aminoácidos não foi identificada em cana-de-açúcar (Figura 22).

Figura 22 - Alinhamento de sequências proteicas de HCT dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

Para PAL, a sequência encontrada no SUCEST que se alinha Sb006G148800 com porção N-terminal de 23 aminoácidos, sendo idêntica a partir de então. Já para Sb004G220300, cada uma das sequências de cana-de-açúcar encontradas consistia em apenas metade do comprimento desta, sendo necessária sobreposição de ambas para que se pudesse obter uma sequência completa (figura 23).

Figura 23 - Alinhamento de sequências proteicas de PAL dos diferentes bancos de cana-de-açúcar com os genes *bona fide* de *S. bicolor*.

(Continua)

Figura 23 - Conclusão

A) Alinhamento com Sb006G148800; B) alinhamento com Sb004G220300. As siglas no início de cada sequência demonstra a fonte dos respectivos bancos transcriptômicos analisados: Sc – SUCEST; Fl – Folha; Rt – Raiz; Pl – Plântula e; En – Entrenó. Fonte: do autor.

Em conjunto, esses alinhamentos mostram diversas sequências com alto grau de identidade e cobertura em relação a sorgo (Tabela 5), sequências incompletas porém com regiões idênticas, e ainda sequências idênticas entre si, provenientes de diferentes bancos de transcriptoma. Esse é um fato interessante, pois mostra que, possivelmente, as mesmas sequências estão envolvidas na lignificação de diferentes tecidos e linhagens de cana-de-açúcar, e permite a seleção de um mesmo gene candidato para futuras análises genéticas.

Baseado na alta similaridade entre as sequências (medido pela identidade do alinhamento) e na proximidade do tamanho entre as sequências de cana-de-açúcar e sorgo (cobertura do alinhamento) foram selecionadas as sequências completas ou próximas desta condição para elaborar as filogenias através do método *Neighbor-joining* dos clados *bona fide* entre as sequências de cana-de-açúcar e gramíneas relacionadas (Figura 24).

Figura 24 - Árvores filogenéticas dos clados *bona fide* das enzimas envolvidas na biossíntese de lignina entre cana-de-açúcar, *Brachypodium*, arroz, milho e sorgo e *Arabidopsis*.

A) PAL; B) 4CL; C) C4H; D) CCR; E) CCoAOMT; F) F5H; G) COMT; H) HCT; I) CAD e; J) C3H.
Fonte: do autor

A partir das análises de bioinformática e evolutiva associado às prévias informações disponíveis na literatura foi possível apontar os melhores candidatos a participar na biossíntese de lignina em cana-de-açúcar (Tabela 5). Sendo estes possíveis ortólogos dos genes *bona fide*, é possível conjecturar que sejam de fato ativos e funcionais em cana-de-açúcar, e assim possuam grande interesse de estudo.

Tabela 5 - Número de genes candidatos em cana-de-açúcar e suas respectivas relações de identidade e cobertura com sequência ortólogo de sorgo.

Enzima	Número de candidatos encontrados	Ortólogo em sorgo	Identidade	Cobertura
4CL	5	Sb007G145600	96%	73%
		Sb010G214900	98%	80%
		Sb004G062500	98%	100%
		Sb007G089900	99%	44%
		Sb004G272700	99%	77%
CCoAOMT	1	Sb010G052200	98%	100%
C3H	2	Sb009G181800	99%	100%
		Sb003G327800	96%	100%
C4H	3	Sb002G126600	97%	100%
		Sb003G337400	98%	68%
		Sb004G141200	98%	86%
F5H	2	Sb001G196300	97%	54%
		Sb005G088400	89%	50%
HCT	2	Sb004G212300	98%	100%
		Sb006G136800	97%	83%
PAL	2	Sb004G220300	97%	74%
		Sb006G148800	98%	100%
CCR	4	Sb007G141200	96%	100%
		Sb004G065600	97%	94%
		Sb002G146000	98%	100%
		Sb010G066000	96%	67%
CAD	1	Sb004G071000	96%	100%
COMT	1	Sb007G047300	95%	100%

Fonte: do autor.

5 CONCLUSÃO

As informações genômicas disponíveis para uma determinada espécie podem ser de grande utilidade no estudo de espécies próximas. A grande quantidade de conhecimento obtido em *Arabidopsis thaliana* pôde ser extrapolada para gramíneas modelo e para cana-de-açúcar, que ainda não possui genoma sequenciado e por isso apresenta maior dificuldade na identificação de sequências de interesse. Com base nas sequências de *Arabidopsis*, bem como milho, sorgo, *brachypodium* e arroz, foi possível identificar genes membros de famílias gênicas que apresentam enzimas envolvidas na via de biossíntese de monolignóis nessas espécies e, a partir de tal, encontrar sequências dessa via em cana-de-açúcar. Por meio da busca por sequências homólogas em diferentes bancos de transcriptoma e subsequente filogenias, foi possível destacar genes candidatos, ortólogos aos *bona fide* das espécies estudadas. Em um total de 23 sequências de cana-de-açúcar, esses candidatos poderão ser estudados buscando relacioná-los às características fenotípicas e à recalcitrância dos materiais lignocelulósicos por meio de análises de expressão diferencial e desenvolvimento de linhagens mutantes.

REFERÊNCIAS

- ACHNINE, L.; BLANCAFLOR, E.B.; RASMUSSEN, S.; DIXON, R.A. Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. **Plant Cell** v16, p3098–3109, 2004.
- ADLER, E. Lignin chemistry: past present and future. **Wood Sci. Technol.** v11, p169–218, 1977.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D. Basic Local Alignment Search Tool. **J. Mol. Biol.**, v215, p403-410, 1990.
- ANDERSON, N.A.; TOBIMATSU, Y.; CIESIELSKI, P.N.; XIMENES, E.; RALPH, J.; DONOHUE, B.S.; LADISCH, M.; CHAPPLE, C. Manipulation of Guaiacyl and Syringyl Monomer Biosynthesis in an Arabidopsis Cinnamyl Alcohol Dehydrogenase Mutant Results in Atypical Lignin Biosynthesis and Modified Cell Wall Structure. **The Plant Cell**, v27, p2195–2209, 2015.
- ARENCIBIA, A.; CARMONA, E.; TELLEZ, P.; CHAN, M.T.; YU, S.M.; TRUJILLO, L.; ORAMAS, P. An efficient protocol for sugarcane (*Saccharum* spp.) transformation mediated by *Agrobacterium tumefaciens*. **Transgenic Res.** v7, p213–222, 1998.
- ARRUDA, P. Genetically modified sugarcane for bioenergy generation. **Curr Opin Biotechnol** v23, p315–322, 2012.
- ATALLA, R.H.; AGARWAL, U.P. Raman microprobe evidence for lignin orientation in the cell walls of native woody tissue. **Science** v227, p636–38, 1985.
- ATALLA, R.H.; BRADY, J.W.; MATTHEWS, J.F.; DING, S-Y.; HIMMEL, M.E. Structures of plant cell wall celluloses. In **Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy**. Ed. Blackwell; p188-212, 2008.
- BASNAYAKE, S.W.V; MORGAN, T.C.; WU, L.; BIRCH, R.G. Field performance of transgenic sugarcane expressing isomaltulose synthase. **Plant Biotech J** v10, p217–225, 2012.
- BATE, N.J.; ORR, J.; NI, W.; MEROMI, A.; NADLER-HASSAR, T.; DOERNER, P.W.; DIXON, R.A.; LAMB, C.J.; ELKIND, Y. Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a

rate-determining step in natural product synthesis. **Proc. Natl. Acad. Sci. USA** v91, p7608–12, 1994.

BAUCHER, M.; CHABBERT, B.; PILATE, G.; VAN DOORSSELAERE, J.; TOLLIER, M.T.; PETIT-CONIL, M.; CORNU, D.; MONTIES, B.; VAN MONTAGU, M.; INZE, D.; JOUANIN, L.; BOERJAN, W. Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar (*Populus tremula* x *P. alba*). **Plant Physiol.** v112, p1479–90, 1996.

BAUCHER, M.; MONTIES, B.; VAN MONTAGU, M.; BOERJAN, W. Biosynthesis and genetic engineering of lignin. **Crit. Rev. Plant Sci.** v17, p125–97, 1998.

BAUCHER, M.; BERNARD-VAILHÉ, M.A.; CHABBERT, B.; BESLE, J.M.; OPSOMER, C.; VAN MONTAGU, M.; BOTTERMAN, J. Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the impact on lignin composition and digestibility. **Plant Mol. Biol.** v39, p437–47, 1999.

BATTLE, M.; BENDER, M.L.; TANS, P.P.; WHITE, J.W.C.; ELLIS, J.T.; CONWAY, T.; FRANCEY, R.J. Global carbon sinks and their variability inferred from atmospheric O₂ and $\delta^{13}\text{C}$. **Science**, v287, p2467-2470, 2000.

BEVAN, M.; UAUY, C. Genomics reveals new landscapes for crop improvement. **Genome Biology**, v14, p206, 2013.

BHUIYAN, N.H.; SELVARAJ, G.; WEI, Y.; KING, J. Role of lignification in plant defense. **Plant Signaling & Behavior** v4:2, p158-159, 2009.

BOERJAN, W.; RALPH, J.; BAUCHER, M. Lignin Biosynthesis. **Annu. Rev. Plant Biol.** v54, p519–46, 2003.

BOLGER, M.E.; WEISSHAAR, B.; SCHOLZ, U.; STEIN, N.; USADEL, B.; MAYER, K.F. Plant genome sequencing-applications for crop improvement. **Curr. Opin. Biotechnol.** v26, p31–37, 2014.

BONAWITZ, N.; CHAPPLE, C. The Genetics of Lignin Biosynthesis: Connecting genotype to Phenotype. **Annu. Rev. Genet.** v44, p337-363, 2010.

BOSCH, M.; MAYER, C.; COOKSON, A.; DONNISON, I.S. Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. **Journal of Experimental Botany**, v62, p3545–3561, 2011.

BOTTCHER, A.; CESARINO, I.; SANTOS, A.B.; VICENTINI, R.; MAYER, J.L.; VANHOLME, R.; MORREEL, K.; GOEMINNE, G.; MOURA, J.C.; NOBILE, P.M.; CARMELLO-GUERREIRO, S.M.; ANJOS, I.A.; CRESTE, S.; BOERJAN, W.; LANDELL, M.G.; MAZZAFERA, P. Lignification in Sugarcane: Biochemical Characterization, Gene Discovery, and Expression Analysis in Two Genotypes Contrasting for Lignin Content. **Plant Physiology**, v163, p1539–1557, 2013.

BOUDET, A.M.; KAJITA, S.; GRIMA-PETTENATI, J.; GOFFNER, D. Lignins and lignocellulosics: a better control of synthesis for new and improved uses. **Trends Plant Sci.** v8, p576-581, 2003.

BREMER, K. Gondwanan evolution of the grass alliance of families (*Poales*). **Evolution** v56, p1374–1387, 2002.

BUCHANAN, B.; GRUISSEM, W.; JONES, R. Biochemistry & molecular biology of plants (1a ed.). **American society of plant physiology**. ISBN: 978-0-943088-39-6, 2000.

BUKH, C.; NORD-LARSEN, P.H.; RASMUSSEN, S.H. Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. **Journal of Experimental Botany**, v63, p695–709, 2012.

BYRT, C.S.; GROF, C.P.L.; FURBANK, R.T. C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. **Journal of Integrative Plant Biology** v53, p120–135, 2011.

CAMPBELL, M.M.; SEDEROFF, R.R. Variation in lignin content and composition. Mechanisms of control and implications for the genetic improvement of plants. **Plant Physiol.** v110, p3–13, 1996.

CARDOSO-SILVA, C.B.; COSTA, E.A.; MANCINI, M.C.; BALSALOBRE, T.W.; CANESIN, L.E.; PINTO, L.R.; CARNEIRO, M.S.; GARCIA, A.A.; DE SOUZA, A.P.; VICENTINI, R. (2014). De Novo Assembly and Transcriptome Analysis of Contrasting Sugarcane Varieties. **PLoS ONE** v9(2), e88462, 2014.

CAROCHA, V.; SOLER, M.; HEFER, C.; CASSAN-WANG, H.; FEVEREIRO, P.; MYBURG, A.A.; PAIVA, J.A.P.; GRIMA-PETTENATI, J. Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. **New Phytologist**. v206, p1297–1313, 2015.

CARPITA, N.C.; MCCANN, M.C. Maize and sorghum: genetic resources for bioenergy grasses. **Trends in Plant Science** V13, p1360-1385, 2008.

CASS, C.L.; PERALDI, A.; DOWD, P.F.; MOTTIAR, Y.; SANTORO, N.; KARLEN, S.D.; BUKHMAN, Y.V.; FOSTER, C.E.; THROWER, N.; BRUNO, L.C.; MOSKVIN, O.V.; JOHNSON, E.T.; WILLHOIT, M.E.; PHUTANE, M.; RALPH, J.; MANSFIELD, S.D.; NICHOLSON, P.; SEDBROOK, J.C. Effects of Phenylalanine Ammonia Lyase (PAL) knockdown on cell wall composition, biomass digestibility, and biotic and abiotic stress responses in *Brachypodium*. **Journal of Experimental Botany**, V66:14, p4317–4335, 2015.

CHASE, M.W.; FAY, M.F.; DEVEY, D.S.; MAURIN, O.; RØNSTED, N.; DAVIES, T.J.; PILLON, Y.; PETERSEN, G.; SEBERG, O.; TAMURA, M.N.; ASMUSSEN, C.B.; HILU, K.; BORSCH, T.; DAVIS, J.I.; STEVENSON, D.W.; PIRES, J.C.; GIVNISH, T.J.; SYTSMA, K.J.; MCPHERSON, M.A.; GRAHAM, S.W. Multigene Analyses Of Monocot Relationships: A Summary. **Aliso** v22, p63–75, 2006.

CHAPPLE, C. Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. **Annu Rev Plant Physiol Plant Mol Biol** v49, p311–343, 1998.

CHEN, C.; MEYERMANS, H.; BURGGRAEVE, B.; DE RYCKE, R.M.; INOUE, K.; DE VLEESSCHAUWER, V.; STEENACKERS, M.; VAN MONTAGU, M.C.; ENGLER, G.J.; BOERJAN, W.A. Cell-specific and conditional expression of caffeoyl-CoA O-methyltransferase in poplar. **Plant Physiol**. v123, p853–67, 2000.

CHEN, F.; DIXON, R.A. Lignin modification improves fermentable sugar yields for biofuel production. **Nat Biotechnol**, v25, p759-761, 2007.

CLAYTON, W.D.; RENVOIZE, S.A. Genera Graminum. Grasses of the world. **Kew Bulletin, Additional Series** v13, p1–389, 1986.

CUADRADO, A.; ACEVEDO, R.; MORENO, S.; JOUVE, N.; TORRE, C. Genome remodeling in three modern *S. officinarum* and *S. spontaneum* sugarcane cultivars. **J. Exp. Bot.** v55, p847–854, 2004.

DAVIN, L.B.; LEWIS, N.G. Phenylpropanoid metabolism: biosynthesis of monolignols, lignans, neolignans, lignins and suberins. **Rec Adv Phytochem**, v26, p325-375, 1992.

DAVIN, L.B., LEWIS, N.G. Lignin primary structures and dirigent sites. **Curr Opin Biotechnol**, v16, p407-415, 2005.

D'HONT, A.; GRIVET, L.; FELDMANN, P.; RAO, S.; BERDING, N.; GLASZMANN, J.C. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Mol. Gen. Genet.** v250, p405–413, 1996.

D'HONT, A.; ISON, D.; ALIX, K.; ROUX, C.; GLASZMANN, J.C. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. **Genome Res.** v41, p221–225, 1998.

DING, S.Y.; HIMMEL, M.E. The maize primary cell wall microfibril: a new model derived from direct visualization. **J Agric Food Chem**, v54, p597-606, 2006.

DIXON, R.A. Natural products and plant disease resistance. **Nature**, v411, p843-847, 2001.

DONALDSON, L.A. Lignification and lignin topochemistry—an ultrastructural view. **Phytochemistry** v57, p859–73, 2001.

DUAN, C.; YU, J.; BAI, J.; ZHU, Z.; WANG, X. Induced defense responses in rice plants against small brown planthopper infestation. **The crop journal**, v2 p55–62, 2014.

EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research** v32(5), p1792-97, 2004.

EDISON, A.; HALL, R.D.; JUNOT, C.; KARP, P.D.; KURLAND, I.J.; MISTRIK, R.; REED, L.K.; SAITO, K.; SALEK, R.M.; STEINBECK, C.; SUMNER, L.W.; VIAN, M.R. The Time Is Right to Focus on Model Organism Metabolomes. **Metabolites**. v15, p6(1), 2016.

EDWARDS, D.; BATLEY, J. Plant genome sequencing: applications for crop improvement. **Plant Biotechnology Journal**, v8, p2–9, 2010.

EHLTING, J.; BÜTTNER, D.; WANG, Q.; DOUGLAS, C.J.; SOMSSICH, I.E.; KOMBRINK, E. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. **Plant J.** v19, p9–20, 1999.

ELKIND, Y.; EDWARDS, R.; MAVANDAD, M.; HEDRICK, S.A.; RIBAK, O.; DIXON, R.A.; LAMB, C.J. Abnormal plant development and down-regulation of phenylpropanoid biosynthesis in transgenic tobacco containing a heterologous phenylalanine ammonia-lyase gene. **Proc. Natl. Acad. Sci. USA** v87, p9057–61, 1990.

EMONS, A.M.C.; MULDER, M. How the deposition of cellulose microfibrils builds cell wall architecture. **Trends in Plant Science**, v5, p35-40, 2000.

ENRIQUEZ-OBREGON, G.A.; VAZQUEZ, P.R.I.; PRIETO, S.D.L.; RIVA-GUSTAVO, A.D.L.; SELMAN, H.G. Herbicide resistant sugarcane (*Saccharum officinarum* L.) plants by *Agrobacterium*-mediated transformation. **Planta** v206, p20–27, 1998.

FENGEL, D.; WEGENER, G. Wood-Chemistry, Ultrastructure, Reactions. Ed: Walter de Gruyter. Berlin, New York, USA, 613pp, 1989.

FINN, R.D.; CLEMENTS, J.; EDDY, S.R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research**, v39, p29–37, 2011.

FINN, R.D.; CLEMENTS, J.; ARNDT, W.; MILLER, B.L.; WHEELER, T.J.; SCHREIBER, F.; BATEMAN, A.; EDDY, S.R. HMMER web server: 2015 update. **Nucleic Acids Research**, v43, p30-38, 2015.

FINN, R.D.; COGGILL, P.; EBERHARDT, R.Y.; EDDY, S.R.; MISTRY, J.; MITCHELL, A.L.; POTTER, S.C.; PUNTA, M.; QURESHI, M.; SANGRADOR-VEGAS, A.; SALAZAR, G.A.; TATE, J.; BATEMAN, A. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Research**, v44, p279-285, 2016.

FORNALÉE, S.; CAPELLADES, M.; ENCINA, A.; WANG, K.; IRAR, S.; LAPIERRE, C.; RUEL, K.; JOSELEAU, J.P.; BERENGUER, J.; PUIGDOMENECH, P.; RIGAU, J.; CAPARROS-RUIZ, D. Altered lignin biosynthesis improves cellulosic bioethanol

production in transgenic maize plants down-regulated for cinnamyl alcohol dehydrogenase. **Molecular Plant**, v5, p817–830, 2012.

FREUDENBERG, K.; NEISH, A.C. Constitution and Biosynthesis of Lignin. Berlin: Springer-Verlag. 129 pp, 1968.

FU, C.; XIAO, X.; XI, Y.; GE, Y.; CHEN, F.; BOUTON, J.; DIXON, R.A.; WANG, Z.Y. Downregulation of Cinnamyl Alcohol Dehydrogenase (CAD) Leads to Improved Saccharification Efficiency in Switchgrass. **Bioenerg. Res.** v4, p153–164, 2011.

GALLEGO-GIRALDO, L.; JIKUMARU, Y.; KAMIYA, Y.; TANG, Y.; DIXON, R.A. Selective lignin downregulation leads to constitutive defense response expression in alfalfa (*Medicago sativa* L.). **New Phytologist**. v190, p627–639, 2011.

GASTEIGER, E.; GATTIKER, A.; HOOGLAND, C.; IVANYI, I.; APPEL, R.D.; BAIROCH, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Res.** v31, p3784-3788, 2003.

GOODSTEIN, D.M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R.D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D.S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Res.** v40, p1178-1186, 2012.

GOUY, M.; GUINDON, S.; GASCUEL, O. SeaView version 4 p. a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Molecular Biology and Evolution**, v27, p221-224, 2010.

GRABBER, J.H.; RALPH, J.; HATFIELD, R.D.; QUIDEAU, S. *p*-Hydroxyphenyl, guaiacyl, and syringyl lignins have similar inhibitory effects on wall degradability. **Journal of Agricultural and Food Chemistry**, v45, p2530–2532, 1997.

GRABBER, J.H. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. **Crop Science**, v45, p820, 2005.

GRASS PHYLOGENY WORKING GROUP II. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. **New Phytologist**, v193, p304–312, 2012.

GREEN, A.R.; LEWIS, K.M.; BARR, J.T.; JONES, J.P.; LU, F.; RALPH, J.; VERMERRIS, W.; SATTLER, S.E.; KANG, C. Determination of the Structure and Catalytic Mechanism of Sorghum bicolor Caffeic Acid O-Methyltransferase and the Structural Impact of Three brown midrib12 Mutations. **Plant Physiology**, v165, p1440–1456, 2014.

GRIFFITH, M.; GRIFFITH, O.L.; MWENIFUMBO, J.; GOYA, R.; MORRISSY, A.S.; MORIN, R.D.; CORBETT, R.; TANG, M.J.; HOU, Y.C.; PUGH, T.J.; ROBERTSON, G.; CHITTARANJAN, S.; ALLY, A.; ASANO, J.K.; CHAN, S.Y.; LI, H.I.; MCDONALD, H.; TEAGUE, K.; ZHAO, Y.; ZENG, T.; DELANEY, A.; HIRST, M.; MORIN, G.B.; JONES, S.J.; TAI, I.T.; MARRA, M.A. Alternative expression analysis by RNA sequencing. **Nat Methods**. v7(10), p843–7, 2010.

GUI, J.; SHEN, J.; LI, L. Functional characterization of evolutionarily divergent 4-coumarate: Coenzyme A ligases in rice. **Plant Physiol**, v157, p574–586, 2011.

GUILLAUMIE, S.; PICHON, M.; MARTINANT, J.P.; BOSIO, M.; GOFFNER, D.; BARRIÈRE, Y. Differential expression of phenylpropanoid and related genes in brown-midrib bm1, bm2, bm3, and bm4 young nearisogenic maize plants. **Planta** v226, p235–250, 2007.

GUO, D.; CHEN, F.; INOUE, K.; BLOUNT, J.W.; DIXON, R.A. Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. **Plant Cell** v13, p73–88, 2001.

HA, S.; MOORE, P.H.; HEINZ, D.; KATO, S.; OHMIDO, N.; FUKUI, K. Quantitative chromosome map of the polyploid *Saccharum spontaneum* by multicolor fluorescence in situ hybridization and imaging methods. **Plant Mol. Biol.** v39, p1165–1173, 1999.

HALPIN, C.; HOLT, K.; CHOJECKI, J.; OLIVER, D.; CHABBERT, B.; MONTIES, B.; EDWARDS, K.; BARAKAT, A.; FOXON, G.A. Brown-midrib maize (bm1)a mutation affecting the cinnamyl alcohol dehydrogenase gene. **Plant J**, v14, p545–553, 1998.

HAHLBROCK, K.; SCHEEL, D. Physiology and molecular biology of phenylpropanoid metabolism. **Annu Rev Plant Physiol Plant Mol Biol**, v40, p347–369, 1989.

HAIGLER, C.H. The relationship between polymerization and crystallization in cellulose biogenesis. In **Biosynthesis and Biodegradation of Cellulose**. Edited by Haigler CH, Weimer P. 1997:99-124, 1997.

HAN, Y.; GAO, S.; MUEGGE, K.; ZHANG, W.; ZHOU, B. Advanced Applications of RNA Sequencing and Challenges. **Bioinformatics and Biology Insights**, v9(S1), p29–46, 2015.

HARDING, S.A.; LESHKEVICH, J.; CHIANG, V.L.; TSAI, C-J. Differential substrate inhibition couples kinetically distinct 4-coumarate: coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. **Plant Physiol.** v128, p428–38, 2002.

HARRISON, M.D.; GEIJSKES, J.; COLEMAN, H.D.; SHAND, K.; KINKEMA, M.; PALUPE, A.; HASSALL, R.; SAINZ, M.; LLOYD, R.; MILES, S.; DALE, J.L. Accumulation of recombinant cellobiohydrolase and endoglucanase in the leaves of mature transgenic sugar cane. **Plant Biotechnol J** v9, p884–896, 2011.

HIMMEL, M.E.; DING, S.; JOHNSON, D.K.; ADNEY, W.S.; NIMLOS, M.R.; BRADY, J.W.; FOUST, T.D. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. **Science**. v315(5813), p804-807, 2007.

HIRANO, K.; AYA, K.; KONDO, M.; OKUNO, A.; MORINAKA, Y.; MATSUOKA, M. OsCAD2 is the major CAD gene responsible for monolignol biosynthesis in rice culm. **Plant Cell Rep**, v31, p91–101, 2012.

HO-YUE-KUANG, S.; ALVARADO, C.; ANTELME, S.; BOUCHET, B.; CÉZARD, L.; BRIS, P.; LEGÉE, F.; MAIA-GRONDARD, A.; YOSHINAGA, A.; SAULNIER, L.; GUILLON, F.; SIBOUT, R.; LAPIERRE, C.; CHATEIGNER-BOUTIN, A.L. Mutation in *Brachypodium* caffeic acid O-methyltransferase 6 alters stem and grain lignins and improves straw saccharification without deteriorating grain quality. **Journal of Experimental Botany**. v67(1), p227-37, 2016.

HU, W-J.; KAWAOKA, A.; TSAI, C-J.; LUNG, J.; OSAKABE, K.; Ebinuma, H.; Chiang, V.L. Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen *Populus tremuloides*. **Proc. Natl. Acad. Sci. USA** v95, p5407–12, 1998.

HU, W.J.; HARDING, S.A.; LUNG, J.; POPKO, J.L.; RALPH, J.; STOKKE, D.D.; TSAI, C.J.; CHIANG, V.L. Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. **Nat. Biotechnol.** v17, p808–12, 1999.

HU, Y.; LI, W.C.; XU, Y.Q.; LI, G.J.; LIAO, Y.; FU, F.L. Differential expression of candidate genes for lignin biosynthesis under drought stress in maize leaves. **J Appl Genet**, v50, p213–223, 2009.

HUMPHREYS, J.M.; CHAPPLE, C. Rewriting the lignin roadmap. **Current Opinion in Plant Biology**, v5, p224–229, 2002.

ILIC, K.; KELLOGG, E.A.; JAISWAL, P.; ZAPATA, F.; STEVENS, P.F.; VINCENT, L.P.; AVRAHAM, S.; REISER, L.; PUJAR, A.; SACHS, M.M.; WHITMAN, N.T.; MCCOUCH, S.R.; SCHAEFFER, M.L.; WARE, D.H.; STEIN, L.D.; RHEE, S.Y. The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant. **Plant Physiology**, v143, p587–599, 2007.

JOUANIN, L.; GOUJON, T.; DE NADAÏ, V.; MARTIN, M.T.; MILA, I.; VALLET, C.; POLLET, B.; YOSHINAGA, A.; CHABBERT, B.; PETIT-CONIL, M.; LAPIERRE, C. Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity. **Plant Physiol.** v123, p1363–73, 2000.

JUNG, J.H.; FOUAD, W.M.; VERMERRIS, W.; GALLO, M.; ALTPETER, F. RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. **Plant Biotechnol J** v10, p1067–1076, 2012.

KAJITA, S.; HISHIYAMA, S.; TOMIMURA, Y.; KATAYAMA, Y.; OMORI, S. Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate: coenzymeA ligase is depressed. **Plant Physiol.** v114, p871–79, 1997.

KEATING, B.A.; WILSON, J.R. Intensive sugarcane production: meeting the challenge beyond 2000. Wallingford: CAB International, 1997.

KELLOGG, E.A. *Brachypodium distachyon* as a Genetic Model System. **Annu Rev Genet.** v49, p1-20, 2015.

KIM, M.; DAY, D. Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. **J Ind Microbiol Biotechnol.** v38(7), p803-7, 2011.

KIM, I.A.; KIM, B.; KIM, M.; AHN, J. Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. **Phytochemistry**, v76, p25–31, 2012.

KINKEMA, M.; GEIJSKES, J.; DELUCCA, P.; PALUPE, A.; SHAND, K.; COLEMAN, H.D.; BRININ, A.; WILLIAMS, B.; SAINZ, M.; DALE, J.L. Improved molecular tools for sugar cane biotechnology. **Plant Mol Biol.** v84(4-5), p497-508, 2014.

KUMAR, M.; CAMPBELL, L.; TURNER, S. Secondary cell walls: biosynthesis and manipulation. **Journal of Experimental Botany**, v67, p515–531, 2016.

LAKSHMANAN, P.; GEIJSKES, R.J.; AITKEN, K.S.; GROF, C.L.P.; BONNETT, G.D.; SMITH, G.R. Sugarcane biotechnology: the challenges and opportunities. **In Vitro Cell Dev Biol Plant** v41, p345–363, 2005.

LAMESCH, P.; BERARDINI TZ, L.D.; SWARBRECK, D.; WILKS, C.; SASIDHARAN, R.; MULLER, R.; DREHER, K.; ALEXANDER, D.L.; GARCIA-HERNANDEZ, M.; KARTHIKEYAN, A.S.; LEE, C.H.; NELSON, W.D.; PLOETZ, L.; SINGH, S.; WENSEL, A.; HUALA, E. The Arabidopsis Information Resource (TAIR) p. improved geneannotation and new tools. **Nucleic Acids Research**, v.40, p.D1202-10, 2011.

LAPIERRE, C.; POLLET, B.; PETIT-CONIL, M.; TOVAL, G.; ROMERO, J.; PILATE, G.; LEPLÉ, J.C.; BOERJAN, W.; FERRET, V.V.; DE NADAI, V.; JOUANIN, L. Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have opposite impact on the efficiency of industrial Kraft pulping. **Plant Physiol.** v119, p153–63, 1999.

LAUVERGEAT, V.; LACOMME, C.; LACOMBE, E.; LASSERRE, E.; ROBY, D.; GRIMA-PETTENATI, J. Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. **Phytochemistry** v57, p1187–95, 2001.

- LEE, D.; MEYER, K.; CHAPPLE, C.; DOUGLAS, C.J. Antisense suppression of 4-coumarate: coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. **Plant Cell** v9, p1985–98, 1997.
- LI, X.; XIMENES, E.; KIM, Y.; SLININGER, M.; MEILAN, R.; LADISCH, M.; CHAPPLE, C. Lignin monomer composition affects Arabidopsis cell-wall degradability after liquid hot water pretreatment. **Biotechnology for Biofuels**, v3, p27, 2010.
- LI, Y.; KIM, J.I.; PYSH, L.; CHAPPLE, C. Four Isoforms of Arabidopsis 4-Coumarate:CoA Ligase Have Overlapping yet Distinct Roles in Phenylpropanoid Metabolism. **Plant Physiology**. v169, p2409–2421, 2015.
- LINDERMAYR, C.; MÖLLERS, B.; FLIEGMANN, J.; UHLMANN, A.; LOTTSPEICH, F.; MEIMBERG, H.; EBEL, J. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. **Eur. J. Biochem.** v269, p1304–15, 2002.
- LIU, N.; LI, M.; GONG, Y.; LIU, F.; LI, T. Cytochrome P450s--Their expression, regulation, and role in insecticide resistance. **Pestic Biochem Physiol.** v120, p77-81, 2015.
- MANICKAVASAGAM, M.; GANAPATHI, A.; ANBAZHAGAN, V.R.; SUDHAKAR, B.; SELVARAJ, N.; VASUDEVAN, A.; KASTHURIENGAN, S. Agrobacterium mediated genetic transformation and development of herbicideresistant sugarcane (*Saccharum* species hybrids) using axillary buds. **Plant Cell Rep.** v23, p134–143, 2004.
- MANSFIELD, S.D.; KANG, K.Y.; CHAPPLE, C. Designed for deconstruction – poplar trees altered in cell wall lignification improve the efficacy of bioethanol production. **New Phytologist**, v194, p91–101, 2012.
- MASARIN, F.; GURPILHARES, D.B.; BAFFA, D.C.F.; BARBOSA, M.H.P.; CARVALHO, W.; FERRAZ, A.; MILAGRES, A.M.F. Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. **Biotechnology for Biofuels** v4, p55, 2011.
- MATHEWS, S.; TSAI, R.C.; KELLOGG, E.A. Phylogenetic Structure In The Grass Family (Poaceae): Evidence From The Nuclear Gene Phytochrome B. **American Journal of Botany** v87(1), p96–107, 2000.

MCGINNIS, S.; MADDEN, T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. **Nucleic Acids Research**, v32, p20-25, 2004.

MEINKE, D.W.; CHERRY, J.M.; DEAN, C.; ROUNSLEY, S.D.; KOORNNEEF, M. *Arabidopsis thaliana*: A model plant for genome analysis. **Science**, v282, p662-682, 1998.

MIEDES, E.; VANHOLME, R.; BOERJAN, W.; MOLINA, A. The role of the secondary cell wall in plant resistance to pathogens. **Frontiers in Plant Science** v5(38), p1-13, 2014.

MOORE, K.J.; JUNG, H.G. Lignin and fiber digestion. **J Range Manage**, v54, p420-430, 2001.

NEEDLEMAN, S.B.; WUNSCH, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **J. Mol. Biol.**, v48, p443-453, 1970.

OOKAWA, T.; INOUE, K.; MATSUOKA, M.; EBITANI, T.; TAKARADA, T.; YAMAMOTO, T.; UEDA, T.; YOKOYAMA, T.; SUGIYAMA, C.; NAKABA, S.; FUNADA, R.; KATO, H.; KANEKATSU, M.; TOYOTA, K.; MOTOBAYASHI, T.; VAZIRZANJANI, M.; TOJO, S.; HIRASAWA, T. Increased lodging resistance in long-culm, low-lignin gh2 rice for improved feed and bioenergy production. **Sci Rep**. v4, p6567, 2014.

OLSEN, K.M.; LEA, U.S.; SLIMESTAD, R.; VERHEUL, M.; LILLO, C. Differential expression of four Arabidopsis PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis. **Journal of Plant Physiology** v165, p1491—1499, 2008.

PAPINI-TERZI, F.S.; ROCHA, F.R.; VÊNCIO, R.Z.; OLIVEIRA, K.C.; FELIX, J.M.; VICENTINI, R.; ROCHA, C.S.; SIMÕES, A.C.; ULIAN, E.C.; DI MAURO, S.M.; DA SILVA, A.M.; PEREIRA, C.A.; MENOSSI, M.; SOUZA, G.M. Transcription profiling of signal transduction-related genes in sugarcane tissues. **DNA Res** v12, p27–38, 2005.

PARIJS, F.R.; RUTTINK, T.; BOERJAN, W.; HAESAERT, G.; BYRNE, S.L.; ASP, T.; ROLDÁN-RUIZ, I.; MUYLLE, H. Clade classification of monolignol biosynthesis gene family members reveals target genes to decrease lignin in *Lolium perenne*. **Plant Biology** v17, p877–892, 2015.

PAULY, M.; KEEGSTRA, K. Cell-wall carbohydrates and their modification as a resource for biofuels. **Plant J**, v54, p559-568, 2008.

PETRASOVITS, L.A.; ZHAO, L.; MCQUALTER, R.B.; SNELL, K.D.; SOMLEVA, M.N.; PATTERSON, N.A.; NIELSEN, L.K.; BRUMBLEY, S.M. Enhanced polyhydroxybutyrate production in transgenic sugarcane. **Plant Biotechnol J** v10, p569–578, 2012.

PINÇON, G.; CHABANNES, M.; LAPIERRE, C.; POLLET, B.; RUEL, K.; JOSELEAU, J.P.; BOUDET, A.M.; LEGRAND, M. Simultaneous down-regulation of caffeic/5-hydroxy ferulic acid-O-methyltransferase I and cinnamoyl-coenzyme A reductase in the progeny from a cross between tobacco lines homozygous for each transgene. Consequences for plant development and lignin synthesis. **Plant Physiol.** v126, p145–55, 2001a.

PINÇON, G.; MAURY, S.; HOFFMANN, L.; GEOFFROY, P.; LAPIERRE, C.; POLLET, B.; LEGRAND, M. Repression of O-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth. **Phytochemistry** v57, p1167–76, 2001b.

PIQUEMAL, J.; LAPIERRE, C.; MYTON, K.; O'CONNELL, A.; SCHUCH, W.; GRIMA-PETTENATI, J.; BOUDET, A.M. Down-regulation in cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. **Plant J.** v13, p71–83, 1998.

POOVAIAH, C.R.; NAGESWARA-RAO, M.; SONEJI, J.R.; BAXTER, H.L.; STEWART, C.N.JR. Altered lignin biosynthesis using biotechnology to improve lignocellulosic biofuel feedstocks. **Plant Biotechnology Journal** v12, p1163–1173, 2014.

RAES, J.; ROHDE, A.; CHRISTENSEN, J.H.; VAN DE PEER, Y.; BOERJAN, W. Genome-Wide Characterization of the Lignification Toolbox in Arabidopsis. **Plant Physiology**, v133, p1051–1071, 2003.

RALPH, J.; HATFIELD, R.D.; PIQUEMAL, J.; YAHIAOUI, N.; PEAN, M.; LAPIERRE, C.; BOUDET, A.M. NMR characterization of altered lignins extracted from tobacco plants downregulated for lignification enzymes cinnamyl-alcohol dehydrogenase and cinnamoyl-CoA reductase. **Proc. Natl. Acad. Sci. USA** v95, p12803–8, 1998.

RALPH, J.; LAPIERRE, C.; MARITA, J.M.; KIM, H.; LU, F.; HATFIELD, R.D.; RALPH, S.; CHAPPLE, C.; FRANKE, R.; HEMM, M.R.; VAN DOORSSELAERE, J.; SEDEROFF, R.R.; O'MALLEY, D.M.; SCOTT, J.T.; MACKAY, J.J.; YAHIAOUI, N.; BOUDET, A.; PEAN, M.; PILATE, G.; JOUANIN, L.; BOERJAN, W. Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. **Phytochemistry** v57, p993–1003, 2001.

RAMOS, R.L.; TOVAR, F.J.; JUNQUEIRA, R.M.; LINO, F.B.; SACHETTO-MARTINS, G. Sugarcane expressed sequences tags (ESTs) encoding enzymes involved in lignin biosynthesis pathways. **Genet Mol Biol** v24, p235–241, 2001.

RASMUSSEN, S.; DIXON, R.A. Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. **Plant Cell**, v11, p1537–1552, 1999.

REDDY, M.S.S.; CHEN, F.; SHADLE, G.; JACKSON, L.; ALJOE, H.; DIXON, R.A. Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). **Proceedings of the National Academy of Sciences of the United States of America**, v102, p16573–16578, 2005.

REIN, P.W. Prospects for the conversion of a sugar mill into a biorefinery. **Proc Inc Soc Sugar Cane Technol** v26, p44–60, 2007.

ROACH, B.T. Origin and improvement of the genetic base of sugarcane. In: Egan, B. T., ed. *Proc. Aust. Soc. Sugarcane Technol.* 11:34–47, 1989.

ROUSSEL, M.R.; LIM, C. Dynamic model of lignin growing in restricted spaces. **Macromolecules** v28, p370–76, 1995.

SAATHOFF, A.J.; SARATH, G.; CHOW, E.K.; DIEN, B.S.; TOBIAS, C.M. Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. **PLoS ONE** v6, pE16416, 2011.

SABALLOS, A.; SATTLER, S.E.; SANCHEZ, E.; FOSTER, T.P.; XIN, Z.; KANG, C.H.; PEDERSEN, J.F.; VERMERRIS, W. Brown midrib2 (Bmr2) encodes the major 4-coumarate: Coenzyme A ligase involved in lignin biosynthesis in sorghum. **Plant J**, v70, p813–830, 2012.

SAKA, S.; GORING, D.A.I. Localization of lignins in wood cell walls. In *Biosynthesis and Biodegradation of Wood Components*. ed. T Higuchi, pp. 51–62. Orlando: Academic. 679 pp, 1985.

SANKOF, D.; KRUSKAL, J.B. Time warps, string edits and macromolecules: The theory and practice of sequence comparison. Addison-Wesley Reading, 1983.

SANTOS, A.B.; BOTTCHEER, A.; KIYOTA, E.; MAYER, J.L.S.; VICENTINI, R.; BRITO, M.S.; CRESTE, S.; LANDELL, M.G.A.; MAZZAFERA, P. Water Stress Alters Lignin Content and Related Gene Expression in Two Sugarcane Genotypes. **J. Agric. Food Chem.** v63, p4708–4720, 2015.

SATTLER, S.E.; SAATHOFF, A.J.; HAAS, E.J.; PALMER, N.A.; FUNNELL-HARRIS, D.L.; SARATH, G.; PEDERSEN, J.F. A Nonsense Mutation in a Cinnamyl Alcohol Dehydrogenase Gene Is Responsible for the Sorghum brown midrib6 Phenotype. **Plant Physiology**, v150, p584–595, 2009.

SATTLER, S.E.; FUNNELL-HARRIS, D.L.; PEDERSEN, J.F. Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. **Plant Sci**, v178, p229–238, 2010.

SATTLER, S.E.; SABALLOS, A.; XIN, Z.; FUNNELL-HARRIS, D.L.; VERMERRIS, W.; PEDERSEN, J.F. Characterization of Novel Sorghum brown midrib Mutants from an EMS-Mutagenized Population. **G3**, v4, p2115-24, 2014.

SHELLER, H.V.; ULVSKOV, P. Hemicelluloses. **Annu Rev Plant Biol** v61, p263–289, 2010.

SELLERS, P.H. On the theory and computation of evolutionary distances. **SIAM J. Appl. Math**, v26, p787-793, 1974.

SETTA, N.; MONTEIRO-VITORELLO, C.B.; METCALFE, C.J.; CRUZ, G.M.; DEL BEM, L.E.; VICENTINI, R.; NOGUEIRA, F.T.; CAMPOS, R.A.; NUNES, S.L.; TURRINI, P.C.; VIEIRA, A.P.; CRUZ, E.A.; CORRÊA, T.C.; HOTTA, C.T.; DE MELLO, V.A.; VAUTRIN, S.; DA TRINDADE, A.S.; VILELA, M.; LEMBKE, C.G.; SATO, P.M.; DE ANDRADE, R.F.; NISHIYAMA, M.Y.JR.; CARDOSO-SILVA, C.B.; SCORTECCI, K.C.; GARCIA, A.A.; CARNEIRO, M.S.; KIM, C.; PATERSON, A.H.;

BERGÈS, H.; D'HONT, A.; DE SOUZA, A.P.; SOUZA, G.M.; VINCENTZ, M.; KITAJIMA, J.P.; VAN SLUYS, M.A. Building the sugarcane genome for biotechnology and identifying evolutionary trends. **BMC Genomics**, v30, p1–17, 2014.

SEWALT, V.J.H.; NI, W.; BLOUNT, J.W.; JUNG, H.G.; MASOUD, S.A.; HOWLES, P.A.; LAMB, C.; DIXON, R.A. Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonialyase or cinnamate 4-hydroxylase. **Plant Physiol.** v115, p41–50, 1997.

SHAFRIN, F.; DAS, S.S.; SANAN-MISHRA, N.; KHAN, H. Artificial miRNA-mediated down-regulation of two monolignoid biosynthetic genes (C3H and F5H) cause reduction in lignin content in jute. **Plant Mol Biol.** v89, p511–527, 2015.

SHEN, H.; MAZAREI, M.; HISANO, H.; ESCAMILLA-TREVINO, L.; FU, C.; PU, Y.; RUDIS, M.R.; TANG, Y.; XIAO, X.; JACKSON, L.; LI, G.; HERNANDEZ, T.; CHEN, F.; RAGAUSKAS, A.J.; STEWART, C.N.; WANG, Z.Y.; DIXON, R.A. A Genomics Approach to Deciphering Lignin Biosynthesis in Switchgrass. **The Plant Cell**, v25, p4342–4361, 2013.

SMITH, R.A.; GONZALES-VIGIL, E.; KARLEN, S.D.; PARK, J.Y.; LU, F.; WILKERSON, C.G.; SAMUELS, L.; RALPH, J.; MANSFIELD, S.D. Engineering Monolignol p-Coumarate Conjugates into Poplar and Arabidopsis Lignins. **Plant Physiology**, v169, p2992–3001, 2015.

SRIVASTAVA, S.; VISHWAKARMA, R.K.; ARAFAT, Y.A.; GUPTA, S.K.; KHAN, B.M. Abiotic stress induces change in Cinnamoyl CoA Reductase (CCR) protein abundance and lignin deposition in developing seedlings of *Leucaena leucocephala*. **Physiol Mol Biol Plants**. v21(2), p197–205, 2015.

STATES, D.J.; GISH, W. Combined use of sequence similarity and codon bias for coding region identification. **J Comput Biol.** v1(1), p39-50, 1994.

STICKLEN, M.B. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. **Nat Rev Genet**, v9, p433-443, 2008.

SYKES, R.W.; GJERSING, E.L.; FOUTZ, K.; ROTTMANN, W.H.; KUHN, S.A.; FOSTER, C.E.; ZIEBELL, A.; TURNER, G.B.; DECKER, S.R.; HINCHEE, M.A.;

DAVIS, M.F. Down-regulation of p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H) and cinnamate 4-hydroxylase (C4H) genes in the lignin biosynthetic pathway of *Eucalyptus urophylla* × *E. grandis* leads to improved sugar release. **Biotechnol Biofuels**, v8, p128, 2015.

TAIZ, L.; ZEIGER E. Fisiologia Vegetal. 3ª ed. Artmed, Porto Alegre. ISBN: 85-363-0291-7, 2004.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution** v30, p2725-2729, 2013.

TAYLOR, J.G.; OWEN, T.P.; KOONCE, L.T.; HAIGLER, C.H. Dispersed lignin in tracheary elements treated with cellulose synthesis inhibitors provides evidence that molecules of the secondary cell wall mediate wall patterning. **Plant J**, v2, p959-970, 1992.

TERASHIMA, N.; FUKUSHIMA, K.; HE, L.F.; TAKABE, K. Comprehensive model of the lignified plant cell wall. In *Forage Cell Wall Structure and Digestibility*, ed. HG Jung, DR Buxton, RD Hatfield, J Ralph, pp. 247–70. Madison: ASACSSA-SSSA. 794 pp, 1993.

TOBIAS, C.M.; CHOW, E.K. Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification. **Planta**, v220, p678–688, 2005.

TORAMAN, H.E.; VANHOLME, R.; BORÉN, E.; VANWONTERGHEM, Y.; DJOKIC, M.R.; YILDIZ, G.; RONSSE, F.; PRINS, W.; BOERJAN, W.; VAN GEEM, K.M.; MARIN, G.B. Potential of genetically engineered hybrid poplar for pyrolytic production of bio-based phenolic compounds. **Bioresource Technology** v207, p229–236, 2016.

TRABUCCO, G.M.; MATOS, D.A.; LEE, S.J.; SAATHOFF, A.J.; PRIEST, H.D.; MOCKLER, T.C.; SARATH, G.; HAZEN, S.P. Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. **BMC Biotechnology**. v13, p61, 2013.

VAN ACKER, R.; VANHOLME, R.; STORME, V.; MORTIMER, J.C.; DUPREE, P.; BOERJAN, W. Lignin biosynthesis perturbations affect secondary cell wall composition

and saccharification yield in *Arabidopsis thaliana*. **Biotechnology for Biofuels**, v6, p46, 2013.

VANHOLME, R.; MORREEL, K.; RALPH, J.; BOERJAN, W. Lignin engineering. **Curr Opin Plant Biol**, v11, p278-285, 2008.

VANHOLME, R.; DEMEDTS, B.; MORREEL, K.; RALPH, J.; BOERJAN, W. Lignin Biosynthesis and Structure. **Plant Physiology**, v153, p895–905, 2010.

VANHOLME, R.; STORME, V.; VANHOLME, B.; SUNDIN, L.; CHRISTENSEN, J.H.; GOEMINNE, G.; HALPIN, C.; ROHDE, A.; MORREEL, K.; BOERJAN, W. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. **Plant Cell**, v24, p3506-3529, 2012.

VANHOLME, R.; CESARINO, I.; RATAJ, K.; XIAO, Y.; SUNDIN, L.; GOEMINNE, G.; KIM, H.; CROSS, J.; MORREEL, K.; ARAUJO, P.; WELSH, L.; HAUSTRAETE, J.; MCCLELLAN, C.; VANHOLME, B.; RALPH, J.; SIMPSON, G.G.; HALPIN, C.; BOERJAN, W. Caffeoyl Shikimate Esterase (CSE) Is an Enzyme in the Lignin Biosynthetic Pathway in *Arabidopsis*. **Science** v341, p1103-1106, 2013.

VANNESTE, K.; MAERE, S.; VAN DE PEER, Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. **Phil. Trans. R. Soc. B**, v369, p1648, 2014.

VARSHNEY, R.K.; RIBAUT, J.M.; BUCKLER, E.S.; TUBEROSA, R.; RAFALSKI, J.A.; LANGRIDGE, P. Can genomics boost productivity of orphan crops? **Nat. Biotechnol.** v30, p1172–1176, 2012.

VERMERRIS, W.; ABRIL, A. Enhancing cellulose utilization for fuels and chemicals by genetic modification of plant cell wall architecture. **Current Opinion in Biotechnology**, v32, p104–112, 2015.

VETTORE, A.L.; DA SILVA, F.R.; KEMPER, E.L.; SOUZA, G.M.; DA SILVA, A.M.; FERRO, M.I.; HENRIQUE-SILVA, F.; GIGLIOTI, E.A.; LEMOS, M.V.; COUTINHO, L.L.; NOBREGA, M.P.; CARRER, H.; FRANÇA, S.C.; BACCI, M.; GOLDMAN, M.H.; GOMES, S.L.; NUNES, L.R.; CAMARGO, L.E.; SIQUEIRA, W.J.; VAN SLUYS, M.A.; THIEMANN, O.H.; KURAMAE, E.E.; SANTELLI, R.V.; MARINO, C.L.; TARGON, M.L.; FERRO, J.A.; SILVEIRA, H.C.; MARINI, D.C.; LEMOS, E.G.; MONTEIRO-

VITORELLO, C.B.; TAMBOR, J.H.; CARRARO, D.M.; ROBERTO, P.G.; MARTINS, V.G.; GOLDMAN, G.H.; DE OLIVEIRA, R.C.; TRUFFI, D.; COLOMBO, C.A.; ROSSI, M.; DE ARAUJO, P.G.; SCULACCIO, S.A.; ANGELLA, A.; LIMA, M.M.; DE ROSA, V.E.; SIVIERO, F.; COSCRATO, V.E.; MACHADO, M.A.; GRIVET, L.; DI MAURO, S.M.; NOBREGA, F.G.; MENCK, C.F.; BRAGA, M.D.; TELLES, G.P.; CARA, F.A.; PEDROSA, G.; MEIDANIS, J.; ARRUDA, P. Analysis and Functional Annotation of an Expressed Sequence Tag Collection for Tropical Crop Sugarcane. **Genome Research**, v12, p2725–2735, 2003.

VICENTINI, R.; BOTTCHE, A.; BRITO, M.S.; DOS SANTOS, A.B.; CRESTE, S.; LANDELL, M.G.; CESARINO, I.; MAZZAFERA, P. Large-Scale Transcriptome Analysis of Two Sugarcane Genotypes Contrasting for Lignin Content. **PlosOne**, v10, pe0137698, 2015.

WANG, J.P.; NAIK, P.P.; CHEN, H.C.; SHI, R.; LIN, C.Y.; LIU, J.; SHUFORD, C.M.; LI, Q.; SUN, Y.H.; TUNLAYA-ANUKIT, S.; WILLIAMS, C.M.; MUDDIMAN, D.C.; DUCOSTE, J.J.; SEDEROFF, R.R.; CHIANG, V.L. Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in *Populus trichocarpa*. **The Plant Cell**, v26, p894–914, 2014.

WANG, M.L.; GOLDSTEIN, C.; SU, W.; MOORE, P.H.; ALBERT, H.H. Production of biologically active GM-CSF in sugarcane: a secure biofactory. **Transgenic Res** v14, p167–178, 2005.

WANG, M.L.; WANG, S.; XIA, G. From genome to gene: a new epoch for wheat research? **Trends in Plant Science**, v20, p380-387, 2015.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet**. v10(1), p57–63, 2009.

WATERMAN, M.S. General methods of sequence comparison. **Math Biol**. v46, v473-500, 1984.

WEI, H.; XU, Q.; TAYLOR, L.E.; BAKER, J.O.; TUCKER, M.P.; DING, S. Natural paradigms of plant cell wall degradation. **Current Opinion in Biotechnology**, v20, p330–338, 2009.

WENG, J.K.; LI, X.; BONAOWITZ, N.D.; CHAPPLE, C. Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. **Curr Opin Biotechnol**, v19, p166-172, 2008.

WIKSTRÖM, N.; SAVOLAINEN, V.; CHASE, M.W. Evolution of the angiosperm: calibrating the family tree. **Proc. C. Soc. Lond.** v268, p2211-2220, 2001.

XU, Z.; ZHANG, D.; HU, J.; ZHOU, X.; YE, X.; REICHEL, K.L.; STEWART, N.; SYRENNE, R.D.; YANG, X.; GAO, P.; SHI, P.; DOEPPKE, C.; SYKES, R.W.; BURRIS, J.N.; BOZELL, J.J.; CHENG, Z.M.; HAYES, D.G.; LABBE, N.; DAVIS, M.; STEWART, C.N.; YUAN, J.S. Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. **BMC Bioinformatics**, v10(Suppl 11): pS3, 2009.

XU, B.; ESCAMILLA-TREVINO, L.L.; SATHITSUKSANO, L.; SHEN, Z.; SHEN, H.; ZHANG, Y.H.P.; DIXON, R.A.; ZHAO, B. Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. **New Phytologist**. v192, p611–625, 2011.

YANG, I.S.; KIM, S. Analysis of Whole Transcriptome Sequencing Data: Workflow and Software. **Genomics Inform**, v13(4), p119-125, 2015.

YIN, Y.; HUANG, J.; GU, X.; BAR-PELED, M.; XU, Y. Evolution of Plant Nucleotide-Sugar Interconversion Enzymes. **PLoS ONE**. v6(11), p 27995, 2011.

YOON, J.; CHO, L.H.; KIM, S.L.; CHOI, H.; KHO, H.J.; AN, G. The BEL1-type homeobox gene SH5 induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. **Plant J**, v79, p717–728, 2014.

YOON, J.; CHOI, C.; AN, G. Roles of lignin biosynthesis and regulatory genes in plant development. **Journal of Integrative Plant Biology**, v57, p 902–912, 2015.

ZENG, Y.; ZHAO, S.; YANG, S.; DING, S. Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. **Current Opinion in Biotechnology**. v27, p38–45, 2014.

ZUBIETA, C.; KOTA, P.; FERRER, J.L.; DIXON, R.A.; NOEL, J.P. Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase. **Plant Cell** v14, v1265–77, 2002.

ANEXO

Anexo A – Tabela das sequências identificadas como envolvidas na via de biossíntese de lignina em cana-de-açúcar.

Enzima	Referência				
	Bottcher et al., 2013		Santos et al., 2015		Vicentini et al., 2015
		Código		Código	*códigos não disponíveis
PAL	ShPAL1	SCCCLR1048D07.g	ShPAL1	SCCCLR194D07.g	PAL 2
		SCJFLR1017B11.g			PAL 3
	ShPAL2	SCEQRT1024E12.g			
	ShPAL3	SCJLRT1013C01.g			
	ShPAL4	SCJFLR1017B11.g			
C4H	ShC4H1	SCCCCL4009H01.g	ShC4H1	SCCCCL4009H01.g	C4H1
		SCJFRZ2030D03.g			C4H2
	ShC4H2	SCRAD1140C06.g			C4H3
4CL	Sh4CL1	SCCCCL3002A03.g			4CL1
	Sh4CL2	SCVPLB1015F12.g	Sh4CL2	SCVPLB1015F12.g	4CL5
	Sh4CL3	SCMCRT2102F02.g			
HCT	ShHCT1	SCCCCL4009E02.g			HCT1
C3H	ShC3H1	SCVPLB6041E07.g	ShC3H1	SCVPLB6041E07.g	C3H1
	ShC3H2	SCQSRT1036E09.g			C3H2
CCoAOMT	ShCCoAOMT1	SCAGCL6012G12.g	ShCCoAOMT1	SCAGCL6012G12.g	CCoAOMT1
		SCCCCLR1069B09.g		SCCCCLR1069B09.g	
		SCCCCLR1079A02.g		SCCCCLR1079A02.g	
		SCRUFL4024C06.b		SCRUFL4024C06.b	
	ShCCoAOMT2	SCJLRT2050C09.g			CCoAOMT2
	ShCCoAOMT3	SCJFRZ2010H06.g	ShCCoAOMT3	SCJFRZ2010H06.g	CCoAOMT3
CCR	ShCCR1	SCCCRZ2C01A04.g	ShCCR1	SCCCRZ2C01A04.g	CCR1
		SCBFRT1064A01.g		SCBFRT1064A01.g	
	ShCCR2	SCCCCL6024F07.g			
		SCVPLB6064C01.g			
COMT	ShF5H1	SCJLRT1022E04.g	ShF5H	SCJLRT1022E04.g	F5H1
	ShCOMT1	SCRFLR1012F12.g	ShCOMT1	SCRFLR1012F12.g	COMT1
		SCJLRT1023B09.g		SCJLRT1023B09.g	COMT2
CAD	ShCAD2	SCBFAM2021E08.g	ShCAD2	SCBFAM2021E08.g	CAD2
		SCEPRZ1011A02.g		SCEPRZ1011A02.g	
	ShCAD8	SCEQLR1029E05.g	ShCAD8	SCEQLR1029E05.g	CAD8
	ShCAD7	SCCCLB1001F10.g			CAD7
	ShCAD6	SCACHR1038E08.g			CAD6

Fonte: do autor.