UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE LORENA

ALINE LARISSA GONÇALVES

**Prospecção e caracterização da família gênica Expansina,
envolvida na modificação estrutural da celulose cristalina em
cana-de-açúcar**

LORENA-SP
2016

ALINE LARISSA GONÇALVES

**Prospecção e caracterização da família gênica Expansina,
envolvida na modificação estrutural da celulose cristalina em
cana-de-açúcar**

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

Orientadora: Prof. Dr. Tatiane da Franca Silva

Versão Original

Lorena-SP

2016

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)

Gonçalves, Aline Larissa

Prospecção e caracterização da família gênica
Expansina, envolvida na modificação estrutural da
celulose cristalina em cana-de-açúcar / Aline
Larissa Gonçalves; orientadora Tatiane da Franca
Silva - Versão Original. - Lorena, 2017.
98 p.

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. 2017
Orientadora: Tatiane da Franca Silva

1. Etanol lignocelulósico. 2. Cana-de-açúcar. 3.
Hidrólise enzimática. 4. Expansina. I. Título. II.
Silva, Tatiane da Franca, orient.

AGRADECIMENTOS

Agradeço primeiramente à Deus por todas as oportunidades de engrandecimento pessoal e profissional que tive no decorrer do meu mestrado e pela fé que me motivou nos momentos difíceis.

Agradeço à minha família e amigos por todo o apoio, pelos momentos em que pude compartilhar minhas alegrias, reclamar minhas mágoas e chorar minhas tristezas. Em cada um destes momentos o amor e a atenção de vocês fizeram toda a diferença.

Agradeço à minha orientadora Tatiane por todos os ensinamentos e aprendizagem que me proporcionou durante estes dois anos de pesquisa, me impelindo a sempre buscar o meu melhor.

Quero agradecer a todos os membros do Laboratório de Biologia Molecular que contribuíram de forma direta e indireta para a realização deste trabalho. Obrigada pelas risadas, pelos momentos de descontração, mas também pelos conselhos e a disponibilidade em me ajudar quando precisei.

Por fim, agradeço aos funcionários do Debiq e da EEL pelo suporte prestado e as agências de fomento CAPES e Fapesp pelo apoio financeiro.

RESUMO

GONÇALVES, A. L. **Prospecção e caracterização da família gênica Expansina, envolvida na modificação estrutural da celulose cristalina em cana-de-açúcar.** 2016. 98 p. Dissertação (Mestrado em Ciências) - Escola de Engenharia de Lorena – Universidade de São Paulo, Lorena-SP, 2017

Com o crescente aumento da demanda energética e a redução dos recursos fósseis, os biocombustíveis obtidos a partir de biomassa lignocelulósica emergem como uma importante fonte de energia alternativa e sustentável. A biomassa lignocelulósica é formada basicamente por celulose, hemicelulose e lignina, cujo agrupamento compõem a complexa matriz da parede vegetal. A celulose é o principal composto de interesse presente na biomassa, uma vez que pode ser quebrada em glicose e usada no metabolismo de microrganismos para a posterior produção de etanol combustível. No entanto, diversos fatores tornam a biomassa recalcitrante ao processo de conversão, o que eleva o custo de produção dos biocombustíveis. Nesse contexto, proteínas aditivas, como as Expansinas vegetais vêm ganhando grande destaque devido à sua capacidade de afrouxar a parede celular por meio do enfraquecimento da ligação entre os polissacarídeos de forma não covalente, o que diminui o estresse da parede celular e facilita assim a quebra da celulose por enzimas específicas. Assim, o presente trabalho teve como objetivo identificar genes da superfamília das Expansinas em quatro espécies de gramíneas (*Zea mays*, *Sorghum bicolor*, *Brachypodium distachyon* e *Saccharum spp.*). As sequências nucleotídicas obtidas a partir de bancos transcriptômicos de cana-de-açúcar foram usadas na construção de árvores filogenéticas, por meio das quais pode-se inferir as relações de ortologia entre espécies e selecionar sequências com potencial aplicação biotecnológica na promoção da sacarificação enzimática, aumento de biomassa e resistência vegetal à estresse abiótico, sendo estes *ShEXPA14*, *ShEXPA24*, *ShEXPB22*, *ShEXPL3*, *ShEXPA1-zm*, *ShEXPA33*, *ShEXPB28*, *ShEXPB3*, *ShEXPB21*, *ShEXPB25* e *ShEXPB4*. Adicionalmente, os genes de sorgo e cana-de-açúcar foram caracterizados quanto aos domínios e motivos conservados das Expansinas de plantas, identificando as diferenças entre as subfamílias que podem contribuir para a maior especificidade das β -expansinas em parede celular de gramíneas.

Palavras-chave: Etanol lignocelulósico. Cana-de-açúcar. Hidrólise enzimática. Expansina

ABSTRACT

GONÇALVES, A. L. **Prospecting and characterization of the Expansin gene family, involved in the structural modification of crystalline cellulose in sugar cane.** 2016. 98 p. Dissertation (Master of Science) - Escola de Engenharia de Lorena – Universidade de São Paulo, Lorena-SP, 2017

With increasing energy demand and the reduction of fossil resources, biofuels obtained from lignocellulosic biomass emerge as an important source of alternative and sustainable energy. The lignocellulosic biomass is basically formed by cellulose, hemicellulose and lignin, whose grouping makes up the complex matrix of the vegetal cell wall. Cellulose is the main compound of interest present in biomass since it can be broken down into glucose and used in the metabolism of microorganisms for the subsequent production of fuel ethanol. However, several factors make the biomass recalcitrant to the conversion process, which raises the biofuel production cost. In this context, additive proteins, such as vegetable Expansins have been gaining prominence due to their ability to loosen the cell wall by weakening the bond between the polysaccharides in a non-covalent way, which decreases the stress of the cell wall and thus facilitates the break of cellulose by specific enzymes. Thus, the present work aimed to identify nucleotide sequences of the Expansinas superfamily in four species of grasses (*Zea mays*, *Sorghum bicolor*, *Brachypodium distachyon* and *Saccharum spp.*). Sugarcane transcripts were used in the construction of phylogenetic trees, through which one can infer the relations of orthology between species and obtain sequences with potential biotechnological application in the promotion of enzymatic saccharification, biomass increase and plant resistance to abiotic stress, these being *ShEXPA14*, *ShEXPA24*, *ShEXPB22*, *ShEXPL3*, *ShEXPA1-zm*, *ShEXPA33*, *ShEXPB28*, *ShEXPB3*, *ShEXPB21*, *ShEXPB25* and *ShEXPB4*. In addition, sorghum and sugarcane genes were characterized by the conserved domains and motifs of plant Expansins, identifying the differences between the subfamilies that may contribute to the greater specificity of the EXPB in the cell wall of grasses.

Keywords: Lignocellulosic ethanol. Sugarcane. Enzymatic hydrolysis. Expansin

LISTA DE FIGURAS

Figura 1: Previsão do aumento da demanda por biocombustíveis até o ano de 2050.....	13
Figura 2: Esquema da formação da parede primária da célula	16
Figura 3: Modelo de ação das Expansinas sobre a parede celular. As microfibrilas de celulose (verde) são ligadas entre si por glicanos (linhas amarelas e vermelhas), que também podem aderir entre si. a) As Expansinas (azul) são direcionadas para a parede celular, onde os glicanos estão tensionados, b) e agem perturbando a ligação entre os glicanos aderidos à superfície da microfibrila, c) deslocando os polímeros da parede e desencadeando o deslizamento no ponto de adesão dos polímeros, permitindo assim o relaxamento da parede celular.....	17
Figura 4: Estrutura predita das proteínas Expansinas de plantas.....	22
Figura 5: Estrutura filogenética do clado EXPB-I	23
Figura 6: Árvore filogenética das sequências proteicas de Expansinas encontradas em <i>Zea mays</i> sp. Destacado em vermelho estão os ramos com os genes de EXPA, em azul EXPB e verde EXPL	37
Figura 7: Árvore filogenética das sequências proteicas de Expansinas encontradas em <i>Sorgum bicolor</i> sp. Destacado em vermelho estão os ramos com os genes de EXPA, em azul EXPB e verde EXPL	38
Figura 8: Árvore filogenética das sequências proteicas de Expansinas encontradas em <i>Brachypodium distachyon</i> sp. Destacado em vermelho estão os ramos com genes de EXPA, em azul EXPB e verde EXPL	38
Figura 9: Árvore filogenética com as sequencias proteicas de EXPA encontradas em arroz (verde), milho (azul), sorgo (vermelho), <i>Brachypodium distachyon</i> (amarelo) e <i>Arabidopsis thaliana</i> (roxo).....	41
Figura 10: Árvore filogenética com as sequencias proteicas de EXPB encontradas em arroz (círculos verdes), milho (círculos azuis), sorgo (círculos vermelhos), <i>Brachypodium distachyon</i> (círculos amarelos) e <i>Arabidopsis thaliana</i> (roxo)	42
Figura 11: Árvore filogenética com as sequencias proteicas de EXPL encontradas em arroz (verde), milho (azul), sorgo (vermelho), <i>Brachypodium distachyon</i> (amarelo) e <i>Arabidopsis thaliana</i> (roxo)	43
Figura 12: Transcritos encontrados nos diferentes bancos analisados. 1- Folhas, entrenós imaturos e intermediários da cv. SP80-;3280 (Nishiyama, <i>et al</i> , 2014); 2- SUCEST (VETTORE <i>et al.</i> , 2001); 3- Folhas jovens da cv. SP70-1143 (Grativol, <i>et al</i> , 2014); 4- Folhas de 6 cv. (Cardoso-Silva, <i>et al</i> , 2014) ; 5- Plântula (Cedido pela prof. Adriana Hemerly-UFRJ); 6- banco de transcritos de diferentes segmentos de folhas da cultivar SP80-3280 (CTBE); 7- banco de transcritos expressos no entrenó dos cultivares IACSP04-065 e IACSP04-627 (Vicentini <i>et. al</i> , 2015)	44
Figura 13: Árvore filogenética da subfamília EXPA com sequências proteicas de cana-de-açúcar (vermelho), sorgo (azul), milho(verde), arroz (roxo) e <i>A. thaliana</i> (azul claro). Destacado pelo quadrado vermelho estão os ramos com sequências de cana homólogas às sequências de arroz com interesse biotecnológico	50

Figura 14: Árvore filogenética da subfamília EXPB com sequências proteicas de cana-de-açúcar (vermelho), sorgo (azul escuro), milho(verde), arroz (roxo) e <i>A. thaliana</i> (azul claro). Destacado pelo quadrado vermelho estão os ramos com sequências de cana homólogas às sequências com interesse biotecnológico	51
Figura 15: Árvore filogenética da família EXPL com sequências proteicas de cana-de-açúcar (vermelho), sorgo (azul), milho(verde); arroz (roxo) e <i>A. thaliana</i> (azul claro). Destacado pelo quadrado vermelho estão os ramos com sequências de cana homólogas às sequências de arroz com interesse biotecnológico	52
Figura 16: Análise de éxons e íntrons em EXPA de cana-de-açúcar	54
Figura 17: Análise de éxons e íntrons em EXPB de cana-de-açúcar	55
Figura 18: Análise de éxons e íntrons em EXPL de cana-de-açúcar	55
Figura 19: Alinhamento múltiplo das sequências de EXPA de cana-de-açúcar (<i>ShEXPA</i>) com sequências de EXPB (<i>S/EXPB</i>) e EXPA (<i>S/EXPA</i>) de tomate. Letras em vermelho destacam a localização dos motivos e resíduos conservados das sequências; setas indicam localização dos domínios e linhas representam a região do peptídeo sinal e as α e β -inserções	58
Figura 20: Alinhamento múltiplo das sequências de EXPB em cana-de-açúcar (<i>ShEXPB</i>) com sequências de EXPA (<i>S/EXPA</i>) e EXPB (<i>S/EXPB</i>) de tomate. Letras em vermelho destacam a localização dos motivos e resíduos conservados das sequências; setas indicam localização dos domínios e linhas representam a região do peptídeo sinal e as α e β -inserções	59
Figura 21: Alinhamento múltiplo das sequências de EXPL e cana-de-açúcar (<i>ShEXPL</i>) com sequências EXPL (<i>S/EXPL</i>) de tomate. Setas indicam as regiões dos domínios D1 e D2	60
Figura 22: Análise de motivos da subfamília EXPA em cana-de-açúcar	61
Figura 23: Análise de motivos da subfamília EXPB em cana-de-açúcar	62
Figura 24: Análise de motivos da subfamília EXPL em cana-de-açúcar	62
Figura 25: Análise dos dados de expressão dos transcritos de cana-de-açúcar em EXPA em diferentes bibliotecas do SUCEST	64
Figura 26: Análise dos dados de expressão dos transcritos de cana-de-açúcar em EXPB em diferentes bibliotecas do SUCEST	64
Figura 27: Análise dos dados de expressão dos transcritos de cana-de-açúcar em EXPL em diferentes bibliotecas do SUCEST	64
Figura 28: A) Predição da estrutura tridimensional de EXPB13 de sorgo. B) Predição da estrutura tridimensional de EXPB13 de cana-de-açúcar. C) Predição da estrutura tridimensional de <i>ShEXPA</i> 1 de cana-de-açúcar. Estruturas em espiral indicam alfa-hélice, fitas representam folhas beta e linhas representam alças de conexão. Regiões em azul marcam resíduos polares das proteínas	67
Figura 29: Comparação estrutural de <i>ShEXPA</i> e <i>ShEXPB</i> 13	68
Figura 30: (A) Eletroforese em gel de agarose das amostras de RNA extraídas dos indivíduos 140.1, 140.2, 89.1 e 89.2 entrenós 1, 3, 5 e 9 e amostras de folha e raiz dos indivíduos 140.1 e 140.2. (B) Diluição das amostras de folha dos indivíduos 140.1 e 140.2	71
Figura 31: Avaliação da especificidade dos primers da subfamília de EXPB de cana-de-açúcar	72

LISTA DE TABELAS

Tabela 1: Quantidade de etanol misturado à gasolina em diferentes países	14
Tabela 2: Número de genes relatados para as quatro subfamílias de Expansinas dos genomas das plantas e para o grupo polifilético EXLX em genomas microbianos selecionados	21
Tabela 3: Código das diferentes amostras de tecidos vegetais analisados quanto a expressão em cana-de-açúcar	33
Tabela 4: Quantidade de genes de Expansinas identificadas em cada subfamília gênica em diferentes espécies de gramíneas	40
Tabela 5: Nomeação dos transcritos EXPA em cana-de-açúcar, correlacionados com o homólogo no genoma de sorgo e milho, o código do fragmento do genoma correspondente e o número de transcritos encontrados nos 7 bancos analisados	46
Tabela 6: Nomeação dos transcritos EXPB em cana-de-açúcar, correlacionados com o homólogo no genoma de sorgo e milho, o código do fragmento do genoma correspondente e o número de transcritos encontrados nos 7 bancos analisados	47
Tabela 7: Nomeação dos transcritos EXPL em cana-de-açúcar, correlacionados com o homólogo no genoma de sorgo e milho, o código do fragmento do genoma correspondente e o número de transcritos encontrados nos 7 bancos analisados	48
Tabela 8: Concentração de RNA total e valor das razões 260/230nm e 260/280nm do córtex em diferentes estágios de amadurecimento, raiz e folha das duplicatas biológicas do híbrido 140	70
Tabela 9: Concentração de RNA total e valor das razões 260/230nm e 260/280nm do córtex em diferentes estágios de amadurecimento das duplicatas biológicas do híbrido 89	70

SUMÁRIO

1- INTRODUÇÃO.....	9
2- REVISÃO BIBLIOGRÁFICA	11
2.1- Cana de açúcar e a produção de bioetanol no brasil	11
2.2- Mecanismo de ação das Expansinas	15
2.3- Família gênica das Expansinas.....	19
2.3.1- Domínios proteicos	19
2.3.2- Evolução da família gênica das Expansinas	22
2.4- Potencialidades de uso biotecnológico das Expansinas	24
2.5- Ferramentas de análise do genoma	25
3-OBJETIVO	27
4- MATERIAL E MÉTODOS	28
4.1- Prospecção em espécies com genoma completo.....	28
4.2- Prospecção de sequências em cana-de-açúcar.....	29
4.3- Análises filogenéticas em cana-de-açúcar	31
4.4- Buscas genômicas	32
4.5- Análise de expressão <i>in silico</i>	32
4.6- Análise da estrutura das sequências proteicas e genômica de Expansinas de cana-de-açúcar.....	33
4.7- Construção do modelo 3D de Expansinas de cana-de-açúcar	34
4.8- Coleta do material vegetal	35
4.9- Desenho dos iniciadores e teste de especificidade por PCR	35
5-RESULTADOS E DISCUSSÃO	37
5.1- Identificação de Expansinas em espécies relacionadas à cana-de-açúcar	37
5.2- Identificação de Expansinas em cana-de-açúcar	43
5.3- Análise da sequência, domínios e motivos conservados da Expansinas.....	53
5.4- Análise de expressão <i>in silico</i> de cana-de-açúcar	63
5.5- Estrutura predita de Expansinas de sorgo e cana-de-açúcar.....	66
5.6- Extração de RNA e validação dos iniciadores para análise de PCR em Tempo Real.....	69
6- CONCLUSÃO	73
7- PERSPECTIVAS	74
REFERÊNCIAS	75
APÊNDICES	81

1 INTRODUÇÃO

A crescente demanda energética dos últimos anos, assim como a instabilidade nos preços do petróleo e o impacto que a queima dos combustíveis fósseis causa no meio ambiente tem levado à busca de fontes alternativas de energia. Nesse contexto, os biocombustíveis têm sido considerados fontes relevantes de energia sustentável, devido à grande quantidade de biomassa disponível e sua condição renovável, além de oferecem segurança energética e reduzem a emissão de gases do efeito estufa (VALENCIA; CARDONA, 2014; SMITH et al., 2013).

Apesar de seus benefícios, a produção primária de biocombustíveis ainda é derivada de culturas usadas para alimentação humana e animal, como milho, soja, cana-de-açúcar, trigo, entre outros. Desta forma, essa fonte de energia, chamada de primeira geração, compete com os recursos alimentares, deslocando a produção e aumentando a insegurança alimentar (SIMS et al., 2010). No entanto, o desenvolvimento de novas tecnologias tornou possível a produção dos biocombustíveis de segunda geração, obtidos a partir de matéria prima lignocelulósica, que estão tornando-se mais atrativos para a indústria do que os de primeira geração. Isso porque é esperado que esses novos combustíveis tenham maior eficiência de aproveitamento da matéria-prima do que os anteriores e também por não competirem diretamente com recursos alimentares (MCCORMICK; HOWARD, 2013).

A biomassa lignocelulósica, obtida principalmente a partir de resíduos agrícolas e madeira é, em geral, composta por três componentes principais: celulose (35-50%), hemicelulose (20-35%) e lignina (10-25%)(OJEDA et al., 2011). No entanto, apesar da biomassa ter um grande potencial para a produção de biocombustíveis e outros produtos, a conversão de suas moléculas de celulose em açúcar é mais difícil do que apenas obtê-lo do caldo da cana. Este tipo de matéria prima é recalcitrante ao processo devido à interação intermolecular dos componentes da parede e ausência de água na estrutura da microfibrila (SANTOS et al., 2012). Com isso, tornam-se necessárias etapas como pré-tratamentos e o emprego de uma grande carga enzimática durante o processo de sacarificação para uma melhor utilização do potencial da biomassa. Entretanto, a adição destas etapas aumentam o custo de produção e, consequentemente, do preço comercial do etanol (LEE et al., 2014).

Nesse contexto, a identificação de proteínas aditivas na etapa de hidrólise enzimática, como as Expansinas, vem ganhando atenção devido à sua capacidade de afrouxar a parede celular através do enfraquecimento da ligação entre os polissacarídeos da parede, o que permite o relaxamento desta estrutura e facilita sua conversão em glicose (COSGROVE, 2000a). Recentes estudos sugerem que as Expansinas não atuam sobre a celulose como um todo, mas sim na interface da matriz, onde as microfibrilas de celulose entram em contato umas com as outras (COSGROVE, 2015). Apesar de alguns aspectos da atuação destas proteínas sobre as microfibrilas de celulose ainda não estarem completamente elucidadas, ocorreram avanços à respeito das relações de estrutura e função, características de ligação, relações evolutivas e da aplicação das Expansinas na desconstrução da biomassa lignocelulósica (LIU; MA; ZHANG, 2015).

Estudos sugerem que as Expansinas podem atuar em sinergismo com celulasas, levando a um maior rendimento no processo de sacarificação (SEKI et al, 2014). Além de sua utilização como aditivos nas etapas de sacarificação, estas proteínas também tem demonstrado atuar sobre o fenótipo das plantas em estudos de superexpressão, desencadeando um aumento do tamanho da planta e, conseqüentemente, proporcionando uma maior biomassa vegetal (ZENONI et al. 2011).

Em conjunto, estes dados tornam as Expansinas proteínas de alto interesse biotecnológico, tanto para a otimização da produção de biocombustíveis quanto para o melhoramento das características fenotípicas da biomassa, sendo necessário maiores estudos quanto ao mecanismo de ação e fatores que influenciam sua atividade.

2 REVISÃO BIBLIOGRÁFICA

2.1 CANA-DE-AÇÚCAR E A PRODUÇÃO DE ETANOL NO BRASIL

A cana-de-açúcar (*Saccharum* spp.) é uma planta nativa do sudeste asiático que consegue ter um bom desenvolvimento em climas tropicais e subtropicais. Há indícios de que as primeiras mudas de cana tenham chegado ao Brasil por volta de 1515 e desde então ocupam um importante papel na economia brasileira. Por meio de seu metabolismo C4, a cana-de-açúcar consegue converter de forma eficiente a radiação solar e CO₂ em moléculas orgânicas que permitem o aumento de sua biomassa, além de possuir uma habilidade única entre a família das *Poaceas* de acumular até 60% do peso seco do seu caule maduro em forma de açúcar (LISBOA et al., 2011).

Em 2010, o Brasil já era o maior produtor mundial de cana-de-açúcar, sendo sua produção concentrada nos estados centro, sudeste e nordeste, com destaque para o estado de São Paulo, que é responsável por grande parte da produção brasileira de cana-de-açúcar, etanol e açúcar comercial. Entretanto, devido ao aumento da demanda energética e o desenvolvimento de tecnologias que permitem a obtenção de produtos de valor agregado, novas áreas de produção de cana-de-açúcar vêm sendo cultivadas em larga escala (OLIVEIRA; MORAES, 2017).

Parte desta expansão deve-se ao fato do grande número de aplicações possíveis oriundas da utilização da cana-de-açúcar. Além do emprego tradicional na produção de açúcar comercial e etanol, o bagaço produzido, antes visto apenas como resíduo usado para a geração de calor e energia, também pode ser utilizado na produção de biocombustíveis e produtos de valor agregado. Estima-se que 1 tonelada de cana-de-açúcar gere 280 kg de bagaço, que é composto por cerca de 50% de celulose, 25% de hemicelulose e 25% de lignina (SINDHU et al., 2016).

O etanol obtido a partir da biomassa lignocelulósica é chamado de etanol de segunda geração e vêm sendo considerados uma parte da solução para a transição dos combustíveis de origem fóssil para fontes de energia renováveis à médio longo prazo (SIMS et al., 2010). Neste contexto, instalações que empregam processos de conversão de biomassa em biocombustíveis, insumos químicos, entre outros vêm ganhando espaço. Estas instalações são chamadas de biorrefinarias e têm seu conceito intimamente relacionado à conversão de

vários tipos de biomassa em produtos de interesse utilizando diferentes tecnologias de processamento (ZHANG; PEI; WANG, 2015).

O etanol feito à base do bagaço de cana-de-açúcar é essencialmente renovável e possui grandes vantagens em relação à gasolina, já que reduz a emissão de gases do efeito estufa e melhora a qualidade do ar nas grandes metrópoles. Além disso, combustíveis fósseis são recursos finitos e de mercado variável, uma vez que grande parte dos maiores produtores de petróleo no mundo possuem uma política instável, o que leva à uma grande oscilação no preço desse produto. Fato que foi evidenciado em 1973, quando ocorreu a primeira crise do petróleo, que desencadeou um aumento dramático do custo desse combustível para países dependentes de importação, como o Brasil (GOLDEMBERG, 2008).

Esse fato levou o governo brasileiro a encorajar o redirecionamento de parte da produção de açúcar para gerar etanol, como uma alternativa ao uso da gasolina, reduzindo assim as importações. Dessa forma, nasceu o Programa Nacional do Álcool, também conhecido como Proálcool, que incentivou a produção de etanol através de subsídios, proporcionando o desenvolvimento desse setor. Em 1980 o preço de produção do etanol era de US\$ 700 para cada 1000 litros, porém, ao longo dos anos, com o surgimento de novas tecnologias e o impulso ocasionado pelas economias de escala houve uma redução do preço, chegando a US\$200 por 1000 litros em 2004, tornando assim, o etanol economicamente competitivo com a gasolina (GOLDEMBERG, 2008).

Esse contexto levou não só o Brasil como diversos outros países a incentivar a produção de biocombustíveis e investir em estudos para o melhoramento das tecnologias e processos de fabricação. Com isso, a demanda mundial de biocombustíveis cresceu ao longo da última década e tende a aumentar rapidamente nos próximos anos (Figura 1 (IEA, 2011; MASUM et al., 2013)). Isso também pode ser evidenciado pelo fato de que no período entre 2007 e 2010 a produção mundial de etanol de primeira geração praticamente dobrou, além disso, com o incentivo do programa *Renewable Fuel Standard (RFS)*, criado pelos Estados Unidos, espera-se que até 2022 se faça uso de 36 bilhões de galões de combustível renovável, sendo que 16 bilhões de galões devem ser obtidos a partir de matéria prima lignocelulósica (DRABIK; CIAIAN; POKRIVČÁK, 2016; SKOLRUD et al., 2016).

Figura 1- Previsão do aumento da demanda por biocombustíveis até o ano de 2050.

Fonte: Adaptado de IEA (2011)

O uso de biocombustível também vem sendo estimulado em muitos países por meio da adição de pequenas porcentagens de etanol em gasolina. No Brasil, a obrigatoriedade da adição de álcool na gasolina vem desde 1976. Esse fato e a expansão do mercado de veículos movidos à álcool e o desenvolvimento do motor *flex*, que aceitam qualquer proporção de mistura de álcool e gasolina, ajudaram na consolidação do mercado desse produto no país. Cerca de 90% dos carros novos que são vendidos no Brasil possuem motor *flex* e a gasolina comercial tem cerca de 20-25% de etanol anidro incorporado. Muitos países já têm implementado ou estão implementando a obrigatoriedade dessa mistura também, sendo a maioria dos veículos movidos à 10% de etanol combinado à gasolina (Tabela 1). Além disso, fabricantes têm procurado desenvolver veículos que funcionem com quantidades maiores de álcool incorporado ao combustível no motor (MASUM et al., 2013).

Tabela 1- Quantidade de etanol misturado à gasolina em diferentes países.

País	Matéria Prima	Porcentagem de etanol misturado na gasolina (%) (v/v)
Brasil	Cana de açúcar	24
USA	Milho	10
Canadá	Milho, trigo, cevada	7.5–10
Colômbia	Cana de açúcar	10
Espanha	Trigo, cevada	–
França	Beterraba sacarina, trigo, milho	–
Suécia	Trigo	5
China	Milho, trigo	–
Índia	Cana de açúcar	5
Tailândia	Mandioca, cana de açúcar, arroz	10

Fonte: Adaptado de (MASUM et al., 2013)

Esta crescente demanda energética necessita de um recurso abundante e eficiente para ser suprida, sendo por isso a biomassa lignocelulósica uma das mais promissoras matérias primas na indústria de etanol. Em geral, a produção de biocombustível a partir desse material inclui cinco etapas principais: pré tratamento, hidrólise, fermentação, purificação e tratamento de efluentes. Em linhas gerais, o processo empregado para gerar etanol consiste na quebra de polissacarídeos da parede, principalmente da celulose, em monossacarídeos fermentescíveis (etapa de sacarificação) e uma vez que a hidrólise está completa, o hidrolisado é fermentado e transformado em álcool (OJEDA et al., 2011).

Contudo, as etapas adicionais, como pré tratamento, são necessárias para a retirada da lignina e redução da cristalinidade da celulose, aumentando a porosidade da biomassa, permitindo assim uma melhor ação das enzimas na conversão da molécula em açúcar (SUN; CHENG, 2002). Diversos tipos de pré tratamentos podem ser empregados na conversão da biomassa, como métodos físicos, físico-químicos, químicos e biológicos. No entanto, o desenvolvimento de inovações nessa etapa é essencial para aumentar a viabilidade comercial desse processo (ASAKAWA et al., 2015).

Atualmente, têm se observado que vários métodos de deslignificação e amorfogênese do material lignocelulósico aumentam a taxa de sacarificação enzimática. Estudos relacionados com proteínas buscam melhorar a produção de enzimas fúngicas, complementando coquetéis enzimáticos com beta-glicosidases ou com substâncias que

auxiliariam a atuação de outras enzimas (LEE et al., 2011). Nesse contexto, pesquisas relacionadas a utilização de proteínas acessórias na sacarificação buscam melhorar o rendimento da produção de etanol, aumentando a taxa de conversão de celulose em açúcar (SEKI et al., 2014).

2.2 MECANISMO DE AÇÃO DAS EXPANSINAS

A parede celular das plantas é uma importante estrutura que mantém a forma da célula, confere força e rigidez, assim como proteção à patógenos e a aderência entre as células. Essa estrutura também contém uma variedade de proteínas e enzimas com importantes papéis na absorção, transporte e secreção de substâncias. Muitas vezes, os tipos celulares são identificados pela estrutura da parede vegetal, demonstrando uma correlação entre a estrutura da parede e a função da célula (RAVEN; EVERT; EICHHORN, 1996).

A celulose é o principal componente da parede celular e é constituída por inúmeros monômeros de glicose. Os polímeros de celulose são agrupados em estruturas chamadas de microfibrilas e devido ao arranjo ordenado de grande parte de suas moléculas, apresentam propriedades cristalinas (RAVEN; EVERT; EICHHORN, 1996). Assim, a parede primária da célula depositada durante o crescimento é uma rede de microfibrilas de celulose embebidas em uma matriz hidrofílica de hemicelulose e pectinas (Figura 2). As microfibrilas de celulose são sintetizadas por um complexo de proteínas da membrana plasmática, formado por celulose sintase, que utilizam como substrato moléculas de UDP-glicose para a formação da celulose (COSGROVE, 2000a).

Figura 2- Esquema da formação da parede primária da célula.

Fonte: (COSGROVE, 2000a)

A parede em crescimento possui uma notável força e flexibilidade, podendo resistir às grandes forças mecânicas exercidas pela pressão de turgência, enquanto que ao mesmo tempo permite que a parede se distenda. Esse processo está sob um rígido controle bioquímico da membrana, que deve produzir parede primária suficiente para permitir a expansão e ao mesmo tempo permanecer adequadamente rígida impedindo o rompimento da célula (RAVEN; EVERT; EICHHORN, 1996).

A extensão da parede requer a perda de sua estrutura, o que acontece por deslizamento ou rearranjo dos polímeros da matriz que reveste as microfibrilas de celulose, um processo mediado, pelo menos em parte, pelas Expansinas. As Expansinas são proteínas com massa molecular em torno de 26 Kb, isoladas primeiramente na parede celular de hipocótilos de pepino e subsequentemente em outros tecidos de plantas (COSGROVE, 2000a).

O conceito de que as Expansinas são a chave do mecanismo endógeno de afrouxamento da parede celular é suportado por quatro evidências principais: Expansinas conseguem induzir, de maneira dependente de pH, o relaxamento do estresse da parede e a extensão desta estrutura isolada (MCQUEEN-MASON; DURACHKO; COSGROVE, 1992); a aplicação dessas proteínas em células vivas estimula o alargamento da parede celular, demonstrando sua atuação *in vivo*; os genes de Expansinas são expressos durante o processo

de alongamento da célula e por último, a redução na expressão destes genes alteram o crescimento da planta (CHO; COSGROVE, 2000).

No início dos estudos com as Expansinas, acreditava-se que essas proteínas possuíam ação hidrolítica. No entanto, Mcqueen-Mason e Cosgrove (1994) propuseram um mecanismo de atuação alternativo, onde estas proteínas enfraqueceriam ligações não covalentes entre os polissacarídeos da parede, permitindo assim a fluência dos polímeros sem hidrolisá-los (Figura 3). Desta forma, as Expansinas atuam na superfície das microfibrilas de celulose, dissociando o polímero localmente da matriz, levando ao movimento das cadeias e relaxamento do estresse da parede. Este modelo mostrou-se o mais condizente com as características da extensão da parede celular mediada por essas proteínas e também consistente com o progressivo enfraquecimento da parede celular pela ação das Expansinas.

Figura 3- Modelo de ação das Expansinas sobre a parede celular. As microfibrilas de celulose (verde) são ligadas entre si por glicanos (linhas amarelas e vermelhas), que também podem aderir entre si. a) As Expansinas (azul) são direcionadas para a parede celular, onde os glicanos estão tensionados, b) e agem perturbando a ligação entre os glicanos aderidos à superfície da microfibrila, c) deslocando os polímeros da parede e desencadeando o deslizamento no ponto de adesão dos polímeros, permitindo assim o relaxamento da parede celular.

Fonte: (COSGROVE, 2000a)

Além da função mais conhecida na extensão da parede celular durante a expansão da célula, as Expansinas também atuam em outros processos de desenvolvimento da planta relacionados com o afrouxamento da parede celular, como amadurecimento de frutas, abscisão de órgãos, germinação de sementes, penetração do tubo polínico (YENNAWAR et al., 2006), formação do xilema (GRAY-MITSUMUNE et al., 2004), formação de micorrizas (associações de tecidos radiculares com fungos simbióticos) (BALESTRINI; COSGROVE; BONFANTE, 2005), desenvolvimento de nódulos de fixação de nitrogênio em leguminosas (GIORDANO; HIRSCH, 2004) e processos de re-hidratação de plantas, que diminuem o

tamanho das células durante a seca e expandem na presença de água (JONES; MCQUEEN-MASON, 2004).

Nesse contexto, Xu *et al.* (2007) identificaram um gene da família das Expansinas em *Agrostis scabra* (AsEXPA1), uma espécie de gramínea resistente à altas temperaturas, regulado positivamente em condições de estresse por calor enquanto seu homólogo é pouco expresso em plantas sensíveis, como por exemplo, *A. stolonifera*. Entretanto, apesar da suposta correlação entre esses fatores, há pouca informação a respeito da regulação dessas proteínas sobre a tolerância ao calor. No entanto, acredita-se que esta resposta envolva mudanças no nível de expressão dos genes Expansinas (XU *et al.*, 2014).

Buscando elucidar esse mecanismo, Xu *et al.* (2014) isolaram e clonaram um gene de Expansina, da espécie de gramínea perene *Poa pratensis* (PpEXP1), para avaliar a tolerância de plantas transgênicas ao calor. A superexpressão de PpEXP1 propiciou um alto conteúdo de clorofila nas folhas, maior quantidade de água armazenada nos tecidos e uma melhor taxa fotossintética em relação ao controle, além de promover a ação de enzimas antioxidantes, resultando no aumento da tolerância do transgênico ao calor.

Resultados similares foram encontrados em plantas de tabaco superexpressando o gene de Expansina TaEXPB23, obtido a partir de trigo (*Triticum spp.*). A superexpressão de TaEXPB23 aumentou a tolerância das plantas à estresse oxidativo, com seu possível envolvimento na regulação da atividade de enzimas antioxidantes, como por exemplo, peroxidases. Esta enzima demonstrou uma maior ligação covalente à parede da célula em plantas transgênicas do que em tabaco selvagem. Também verificou-se que as plantas transgênicas apresentaram maior quantidade de cloroplasto nas folhas do que o controle e menos sinais de danos por estresse oxidativo (HAN *et al.*, 2015).

Com base em trabalhos anteriores, Han *et al.* (2014) hipotetizou que TaEXPB23 também poderia desempenhar um importante papel adaptativo às mudanças na quantidade de fósforo no solo. Através de seus resultados, os autores concluem que a superexpressão dessa proteína estaria envolvida no aumento da disponibilidade de fósforo por meio da regulação e desenvolvimento da raiz (HAN *et al.*, 2014).

As Expansinas ainda possuem função conhecida no mecanismo de afrouxamento da parede celular, onde permitem o crescimento e expansão das células, alterando o fenótipo vegetal e produzindo maior biomassa (ZENONI *et al.*, 2004). Em seu trabalho, Zenoni *et al.*

(2011) observou que a superexpressão de *PhEXPA1* em plantas de *Petunia hybrida* resultou em mudanças morfológicas no tamanho da pétala, entrenós e área foliar.

Estes dados indicam que proteínas Expansinas desempenham um papel importante no desenvolvimento vegetal. No entanto, devido ao tamanho da superfamília gênica à qual pertencem, a caracterização da função e expressão de muitos genes permanece desconhecida. Embora a expressão de alguns genes específicos tenha demonstrado indução por hormônios, estresse hídrico e outros fatores abióticos, a via de sinalização ainda não foi estudada em detalhes (COSGROVE, 2015; HAN et al., 2012).

2.3 FAMÍLIA GÊNICA DAS EXPANSINAS

A superfamília gênica que codificam as proteínas Expansinas em plantas é normalmente dividida em quatro subfamílias: α -Expansina (*EXPA*), β -Expansina (*EXPB*), Expansina-like A (*EXLA*) e Expansina-like B (*EXLB*) (ZHANG et al., 2014a). Estudos demonstraram experimentalmente que EXPA e EXPB atuam no afrouxamento da parede celular, permitindo seu crescimento (COSGROVE; BEDINGER; DURACHKO, 1997), já as Expansinas-like, tanto A como B, constitui um clado separado dos demais, porém bem estabelecido. Por causa dos peptídeos sinais que os direcionam para a parede celular e por sua estrutura predita ser similar às Expansinas já estudadas, eles provavelmente também atuam na modificação da parede celular (COSGROVE, 2015)

2.3.1 DOMÍNIO PROTEICOS

Proteínas similares às Expansinas encontradas em vegetais também estão presentes em fungos, bactérias e algas (Tabela 2), mas possuem uma relação distante com as proteínas de plantas, o que pode ser constatado através da ocorrência de muitas inserções e deleções observados em alinhamentos (COSGROVE, 2015). Em sua grande maioria, Expansinas de plantas e Expansinas-like de fungos, também chamadas *swolleninas*, assim como as de outros

microrganismos possuem dois domínios compactos chamados de forma geral de D1 e D2 (KERFF et al., 2008).

O domínio D1, comum à Expansinas-like e Expansinas vegetais, é um domínio catalítico que se assemelha à família glicosil hidrolase 45 (GH45). No entanto, ao contrário destas não possuem atividade enzimática, o que pode ser devido à perda de resíduos essenciais para essa função (LIU et al., 2014a). D1 é um domínio localizado na região N-terminal da proteína formado por uma cadeia de seis *double-psi beta-barrel* (DPBB), firmemente acoplado à um domínio de posição C-terminal que acredita se agir como módulo de ligação (domínio D2), tendo uma folha β -pregueada entre eles (GEORGELIS; YENNAWAR; COSGROVE, 2012).

O domínio D2, é indicado como o responsável pela ligação ao substrato, sendo classificado como similar ao "módulo de ligação à carboidratos" - CBM 63, tendo resíduos aromáticos e polares que formam uma potencial superfície de ligação para polissacarídeos (GEORGELIS; YENNAWAR; COSGROVE, 2012). D2 também é homólogo à domínios conhecidos como pólen alergênicos (YENNAWAR et al., 2006).

Tabela 2- Número de genes relatados para as quatro subfamílias de Expansinas dos genomas das plantas e para o grupo polifilético EXLX em genomas microbianos selecionados.

Espécies	EXPA	EXPB	EXLA	EXLB	Total
<i>Arabidopsis thaliana</i> (SAMPEDRO et al., 2005)	26	6	3	1	36
Álamo (SAMPEDRO; CAREY; COSGROVE, 2006)	27	3	2	4	36
Uva (DAL SANTO et al., 2013)	20	4	1	4	29
Soja (ZHU et al., 2014)	49	9	2	15	75
<i>Medicago truncatula</i> (ZHU et al., 2014)	16	1		1	18
<i>Phaseolus vulgaris</i> (ZHU et al., 2014)	25	6		5	36
Maça (ZHANG et al., 2014)	34	1	2	4	41
Repolho chinês (KRISHNAMURTHY et al., 2014)	39	9	2	3	53
Arroz (SAMPEDRO et al., 2005)	33	18	4	1	56
Milho (ZHANG et al., 2014c)	36	48	4		88
Plantas sem floração					
<i>Selaginella moellendorffii</i> (CAREY; HEPLER; COSGROVE, 2013)	15	2			17
<i>Physcomitrella patens</i> (CAREY; HEPLER; COSGROVE, 2013)	28	7			34
Algas					'EXP*'
<i>Micrasterias denticulata</i> (VANNERUM et al., 2011)					4
Bactérias e fungos (NIKOLAIDIS; DORAN; COSGROVE, 2014)					EXLX*
<i>Bacillus subtilis</i>					1
<i>Clavibacter michiganensis</i>					2
<i>Ralstonia solanacearum</i>					1
<i>Xanthomonas campestris</i>					1
<i>Aspergillus Níger</i>					1
<i>Sclerotinia sclerotiorum</i>					1
<i>Magnaporthe oryzae</i>					1
<i>Trichoderma reesei</i>					3

*ELXL: Nome dado às Expansinas de bactérias e fungos

Fonte: Adaptado de (COSGROVE, 2015)

Em seu estudo, Cosgrove (2000) observou que a semelhança de D1 de Expansinas de plantas com o domínio catalítico das GH45 é cerca de 30% de identidade, enquanto D2 pode apresentar até 50% de identidade com as proteínas encontradas em grãos de pólen de gramíneas (Figura 4), demonstrando uma possível origem e diversificação desses domínios. Estes dados fornecem informações valiosas sobre o mecanismo de ação das Expansinas sobre

as microfibrilas de celulose, elucidando, pelo menos em parte, a influência dos domínios na atividade da proteína (LIU; MA; ZHANG, 2015).

Figura 4- Estrutura predita das proteínas Expansinas de plantas.

Fonte: Adaptado de (COSGROVE, 2000a)

2.3.2 EVOLUÇÃO DA FAMÍLIA GÊNICA DAS EXPANSINAS

Não se sabe ao certo o processo de evolução dos genes Expansinas, mas acredita-se que as subfamílias *EXPA* e *EXPB* já existiam quando as plantas vasculares divergiram dos musgos (LI et al., 2002), já as Expansinas-like *A* e *B* não há indícios de que estivessem presentes além do último ancestral comum das gimnospermas e angiospermas (COSGROVE; BEDINGER; DURACHKO, 1997).

As Expansinas α e β são de tamanho semelhante e mantêm certo número de resíduos conservados e motivos característicos. Estas proteínas apresentam 20-40 % de identidade, o que indica que *EXPA* e *EXPB* possuem funções similares, mas atuam sobre diferentes componentes da parede celular (YENNAWAR et al., 2006). As *EXPB*, entretanto, diferem das *EXPA* em importantes pontos: possuem uma fraca ligação para a parede de eudicotiledôneas (Tipo 1) e uma forte ligação para a parede de gramíneas (Tipo 2). O contrário pode ser observado para as *EXPA*, que atuam melhor sobre parede de eudicotiledôneas do que de gramíneas. Essa seletividade sugere que as *EXPBs* agem sobre estruturas ou polímeros específicos da parede de gramíneas, sendo alguns possíveis alvos moléculas de glucuronarbiloxilanos e β -1-3,1-4 glucano (CARPITA, 1996; COSGROVE, 2000b).

As α -expansinas mostraram se altamente conservadas durante a evolução das plantas terrestres (LEE; CHOI; KENDE, 2001). Diversos estudos demonstram a atividade de *EXPA* no

afrouxamento da parede, notando que seu nível de expressão gênica é maior quando a célula está em crescimento ou diferenciação e particularmente alto em células espessas. Assim, é possível estabelecer uma correlação próxima entre o crescimento da planta e a expressão de α -Expansinas (COSGROVE, 2000a).

Já em EXPB, uma pequena identidade entre as sequências do grupo 1 de pólen alergênicos foi encontrado após o sequenciamento do gene dessas proteínas. No entanto, genes de β -Expansinas são expressos em uma variedade de outros tecidos de plantas e, em geral, perderam epítopes alergênicos específicos encontrados nas proteínas do grão de pólen (LI et al., 2003).

Sampedro *et al.*(2015) demonstraram em seus estudos que houve uma grande irradiação de genes *EXPB* em gramíneas, além de uma ação específica para a conformação da parede celular desse grupo. Os autores propõem que o clado EXPB-I divergiu em gramíneas. Na figura 5, pode-se observar a estrutura filogenética do clado EXPB-I, que consiste em um ramo chamado conservativo, formado por EXPBs de eudicotiledôneas e monocotiledôneas, incluindo as gramíneas e um ramo divergente em monocotiledôneas encontradas apenas em gramíneas, com clados de EXPBs expressas preferencialmente em grãos de pólen, (indicado pela seta em vermelho) e em tecidos vegetativos.

Figura 5- Estrutura filogenética do clado EXPB-1.

Fonte: Adaptado de (COSGROVE, 2015)

2.4 POTENCIALIDADES DE USO BIOTECNOLÓGICO DAS EXPANSINAS

Com o grande número de genes encontrado na superfamília das Expansinas foi possível notar diferenças no padrão de expressão e função entre as proteínas, variando entre células e órgãos específicos, assim como em resposta às condições externas e à hormônios (SAMPEDRO; COSGROVE, 2005a). Estudos demonstram que algumas Expansinas podem melhorar a eficiência da sacarificação enzimática o que proporcionariam um aumento da competitividade dos preços do etanol de segunda geração (LEE et al., 2014).

Em suas análises Seki *et al.* (2014) observaram que três Expansinas diferentes de arroz (*OsEXPA33*; *OsEXPB8*; *OsEXLA3*) foram capazes de auxiliar a degradação de celulose cristalina durante a etapa de sacarificação enzimática. Por meio dos resultados obtidos, os autores observaram que a quantidade de glicose produzida cresce com o aumento da concentração de Expansinas no processo, mesmo em baixas concentrações de celulasas. Isso ocorre devido a ação sinérgica das Expansinas com as celulasas, o que pôde ser observada pela maior taxa de conversão de açúcar quando ambas as proteínas foram adicionadas.

Com base nos resultados obtidos, foi proposto um modelo de ação, onde as Expansinas se ligariam primeiramente às regiões mais acessíveis da celulose, que seria rapidamente hidrolisada pelas celulasas. Logo, uma nova região acessível seria exposta novamente à ação das Expansinas, desestabilizando as ligações da molécula de celulose, permitindo a ação das celulasas nas regiões antes inacessíveis (SEKI et al., 2014).

O trabalho de Seki *et al.* (2014) mostrou se relevante para as pesquisas com Expansinas, uma vez que poucos estudos reportam o sinergismo entre essas proteínas e as celulasas devido à dificuldade de expressá-las de forma heteróloga. As proteínas vegetais, muitas vezes, formam corpos de inclusão dentro das células de bactérias hospedeiras, tornando difícil a purificação e obtenção das mesmas (LIU et al., 2014b). No entanto, o trabalho citado teve sucesso em expressar Expansinas vegetais de forma heteróloga em *E. coli*, assim como o trabalho de Liu *et al.* (2014b), que conseguiu expressar Expansinas em levedura (*P. Pastoris*), demonstrando a viabilidade desse processo para obtenção destas proteínas. Os dados sugerem a possibilidade de utilização destas proteínas como aditivo aos coquetéis enzimáticos, melhorando a eficiência na taxa de sacarificação. Uma vez que as enzimas são um dos maiores gastos do processo de conversão de biomassa, a otimização da

etapa de sacarificação poderia proporcionar um aumento da competitividade dos preços do bioetanol obtidos a partir de matéria prima lignocelulósica (SEKI et al., 2014).

Enquanto alguns genes que codificam Expansinas podem ser utilizados para o aumento do rendimento da sacarificação outros podem agir em condições de estresse abiótico ou aumento de biomassa, como previamente descrito no item 2.2 (HAN et al., 2015; XU et al., 2014; ZENONI et al., 2011). Portanto, a manipulação de genes de Expansinas apresentam se como uma estratégia de potencialização da formação de biomassa vegetal, mediante o mecanismo de controle do crescimento da planta e morfogênese (XU et al., 2013) e otimização da hidrólise enzimática .

Apesar da sua importância biotecnológica e nos diversos processos vegetais envolvidos, o mecanismo de atuação dessas proteínas, assim como interações entre as moléculas, ainda não está precisamente elucidada (COSGROVE, 2015; MA et al., 2013). Com base na atual evolução tecnológica, também aplicada ao desenvolvimento da pesquisa, uma rápida identificação de genes ocorre por meio de ferramentas de bioinformática e métodos estatísticos modernos, o que pode facilitar os trabalhos de identificação e caracterização das Expansinas (LIU; MA; ZHANG, 2015).

2.5 FERRAMENTAS DE ANÁLISE DO GENOMA

A bioinformática associa diversas áreas de conhecimento a fim de estudar o código genético por meio de modelos lógico-matemáticos e estatísticos. Diversos programas desenvolvidos para interpretação de dados biológicos permitem que pesquisadores moleculares possam empregar ferramentas computacionais na elucidação e interpretação de eventos biológicos, através de técnicas modernas de sequenciamento de genes e proteínas (ARAÚJO et al., 2008).

Dentre as ferramentas de bioinformática, o BLAST (*Basic Local Alignment Search Tool*) é um dos programas mais amplamente utilizados. Esse programa faz uma comparação entre as sequências de interesse com as sequências disponíveis em banco de dados (genomas, transcritos e proteínas) em sites específicos, ou de forma local. Outra importante ferramenta de análise do genoma é a chamada tecnologia NGS (*Next Generation Sequencing*), que permite o sequenciamento de genomas ou transcritos, gerando assim dados à serem analisados. Essa

técnica é um poderoso instrumento de obtenção de dados genômicos e transcriptômicos de organismos que não são modelos ou que ainda não possuem seu genoma completamente sequenciado, porém requer uma análise avançada em bioinformática (CHIARA; HORNER; SPADA, 2013).

Apesar da cana-de-açúcar ser uma das mais relevantes plantas de cultivo em clima tropical, a complexidade de seu genoma poliplóide torna seu sequenciamento bastante difícil. Desta forma, buscando iniciar o trabalho de sequenciamento da cana, em setembro de 1988 foi lançado o projeto SUCEST (*Sugarcane expressed sequence tag*) com o objetivo de sequenciar de forma aleatória ESTs (*Expressed Sequence Tag*) e identificar cerca de 50.000 ESTs exclusivos (VETTORE et al., 2001).

A partir do SUCEST, novos trabalhos de sequenciamento foram iniciados, originando bancos de dados transcriptômicos que permitiram a ampliação do estudo genético da cana-de-açúcar. Atualmente, o avanço na área molecular e a disponibilidade de novas técnicas também possibilitaram o desenvolvimento de projetos de sequenciamento do genoma da cana, como o realizado pelo Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), já com dados parciais disponíveis para pesquisas. Com isso, a junção do conjunto de ferramentas disponíveis, como programas de bioinformática e os dados obtidos a partir de bibliotecas de ESTs e de RNAseq, permitem que uma análise mais profunda dos genes de cana seja realizada, possibilitando a identificação de sequências de interesse por meio de proteínas homólogas que compartilham características funcionais (VETTORE et al., 2003).

Atualmente diversos programas de bioinformática são capazes de analisar filogeneticamente sequências de nucleotídeos ou aminoácidos de espécies diferentes por meio de suas relações evolutivas, o que permite construir árvores filogenéticas indicando genes ou proteínas relacionadas evolutivamente entre si. A partir destas relações é possível encontrar proteínas homólogas que podem apresentar uma pressão evolutiva entre si para manter sua função preservada (RIZZO; ROUCHKA, 2007; PEI; GRISHIN, 2004).

Desta forma, o uso destas ferramentas permite avanços na identificação, estudo de função e mecanismo de ação das Expansinas e possibilitarão o aproveitamento do seu potencial para otimizar processos industriais e no melhoramento de cana-de-açúcar, com foco na produção de biocombustíveis a partir da parede celular.

3 OBJETIVOS

O estudo proposto teve como objetivo a identificação e caracterização de genes que compõem a família multigênica das Expansinas em cana-de-açúcar e em espécies de gramíneas relacionadas (*Zea mays*, *Sorghum bicolor* e *Brachypodium distachyon*), buscando, assim, encontrar possíveis genes de interesse que possam atuar na eficiência da sacarificação do bagaço.

Dentro desse objetivo geral pode-se destacar alguns objetivos específicos:

- Prospecção de genes das subfamílias de α , β e Expansina-like nos bancos de transcriptomas e genoma parcial de cana-de-açúcar, assim como no genoma de espécies filogeneticamente relacionadas.
- Realização de análise evolutiva e caracterização das subfamílias gênicas.
- Caracterização da estrutura dos genes de Expansinas e das proteínas codificadas.
- Análise de expressão *in silico* de transcritos de cana-de-açúcar
- Predição de estrutura 3D de proteínas Expansinas

4 MATERIAL E MÉTODOS

4.1 PROSPECÇÃO EM ESPÉCIES COM GENOMA COMPLETO

Antes de iniciar a etapa de identificação de Expansinas em cana-de-açúcar, primeiramente fez-se necessário a prospecção dessas proteínas em plantas relacionadas que já possuíam seu genoma sequenciado. A utilização de proteínas homólogas de espécies relacionadas auxilia em uma identificação mais completa de genes e transcritos em cana-de-açúcar. A escolha das espécies utilizadas no presente estudo foi baseada principalmente na proximidade filogenética com a planta de interesse. Desta forma, foram selecionadas inicialmente três gramíneas: *Oryza sativa* (arroz), *Zea mays* (milho) e *Sorghum bicolor* (sorgo).

Entre essas espécies estudadas, arroz tem grande importância como organismo modelo para as monocotiledôneas, além de possuir sequências de Expansinas já previamente caracterizadas (CHO; KENDE, 1997). Milho foi utilizado devido a sua proximidade com cana-de-açúcar e assim como arroz, já possui Expansinas caracterizadas pelo trabalho de Zhang *et al.* (2014). Por outro lado, a escolha da análise em sorgo foi estrategicamente proposta pelo alto nível de conservação gênica e regiões cromossômicas em sintenia, o que tem sido utilizado como ferramenta no estudo de genética comparativa e identificação de genes entre diferentes espécies de gramíneas (JAKOB; ZHOU; PATERSON, 2009).

Posteriormente, também foi adicionado ao estudo a espécie *Brachypodium distachyon*. Essa gramínea possui seu genoma completamente sequenciado, além de pequeno porte e padrão de crescimento relativamente rápido. Essas características proporcionam a possibilidade da utilização dessa espécie como planta modelo por sua facilidade de manuseio e maior facilidade de transformação genética (VAIN, 2011), o que pode auxiliar em estudos posteriores sobre o mecanismo de ação das Expansinas em gramíneas.

Para o início da prospecção dos genes de Expansinas nas diferentes espécies, primeiramente foi necessário a busca por sequências proteicas já descritas como pertencentes à essa família. Assim, foi possível encontrar sequências de proteínas caracterizadas como Expansinas em *Oryza sativa*, disponíveis no site www.personal.psu.edu/fsl/ExpCentral. As sequências proteicas de arroz foram usadas como base para a procura de proteínas em sorgo, *Brachypodium* e milho no genoma destas gramíneas no banco de dados do PHYTOZOME (<http://www.phytozome.net>), através do

algoritmo BLASTp disponibilizado pelo próprio site. O número de acesso correspondente a cada gene pode ser encontrado no apêndice A.

Para averiguar se as sequências proteicas retornadas no BLAST pertenciam à superfamília das Expansinas, estabeleceu-se um critério baseado na estrutura da proteína. Assim, para as proteínas candidatas seguirem para as análises subsequentes tornou-se necessário detectar a presença dos dois domínios funcionais encontrados em Expansinas de plantas: DPBB e *Pollen Allergen*. Adicionalmente, foi verificada a presença putativa do peptídeo sinal de endereçamento proteico para a via de secreção. A identificação destas características foi feita por meio de programas de bioinformática, sendo os domínios confirmados pelo banco de dados do Pfam (<http://pfam.xfam.org/>) e a presença de peptídeo sinal por meio do programa SignalP (<http://www.cbs.dtu.dk/services/SignalP/>).

Após a confirmação dos genes de Expansinas, as sequências nucleotídicas e proteicas de sorgo e *Brachypodium* foram nomeados de acordo com sua posição e o número do cromossomo no qual foi encontrado. No entanto, as sequências de milho não seguiram essa definição pois já haviam sido descritos e nomeadas em um artigo recentemente publicado por Zhang *et al.* (2014), sendo a nomenclatura dada por eles mantida nesse trabalho.

Em seguida, realizou-se análises filogenéticas, baseado em uma matriz de distância entre as sequências proteicas identificadas, empregando o programa MEGA 6 (TAMURA *et al.*, 2013). As sequências obtidas em cada espécie foram alinhadas pelo próprio programa, através da ferramenta do *Muscle*, sendo posteriormente empregado o método *Neighbor Joining*, nos seguintes parâmetros: *pairwise deletion*, *p-distance* e 1000 repetições de *bootstrap* para a construção das árvores.

4.2 PROSPECÇÃO DE SEQUÊNCIAS EM CANA-DE-AÇÚCAR

Após a identificação e classificação das subfamílias gênicas de Expansinas nas espécies relacionadas à cana-de-açúcar, deu-se início a etapa subsequente do trabalho. Para isso, as sequências nucleotídicas de sorgo foram usadas como moldes na realização de um BLAST local, termo referido para análise que não usa ferramentas disponíveis na internet, sendo a busca feita por um programa no próprio computador. Para a validação dos transcritos obtidos

por esse método, também foi utilizado o programa HMMER (EDDY, 2009), que é capaz de encontrar genes homólogos entre os dados fornecidos com base em métodos de interferência probabilística e modelos de Markov. Por meio dessas análises foi possível identificar sequências nucleotídicas candidatas à genes Expansinas utilizando sete bancos de transcritos de cana-de-açúcar.

Dentre as bases de dados analisadas, seis já estão atualmente disponíveis na literatura, sendo estes: o SUCEST, constituído por transcritos de diferentes órgãos vegetais de cana-de-açúcar (VETTORE et al., 2001); o banco de ESTs (*Expressed Sequence Tag*) de folhas, entrenós imaturos e intermediários da cultivar SP80-3280, publicado por Nishiyama *et al.* (2014), o conjunto de RNAseq de amostras de folhas obtidas a partir de seis tipos de cultivares diferentes (IACSP95-3018; IACSP93-3046; SP81-3250; SP80-3280; RB925345 e RB835486) (CARDOSO-SILVA et al., 2014), ESTs de folhas jovens da cultivar SP70-1143 (Grativol, *et al.*, 2014); o banco de transcritos de folha do cultivar SP80-3280 disponível no site <http://bce.bioetanol.cnpem.br> pelos pesquisadores do Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE) e por fim o banco de transcritos expressos no entrenó dos cultivares IACSP04-065 e IACSP04-627 (VICENTINI et al., 2015). O sétimo banco foi cedido gentilmente pela Professora Adriana Hemerly, da Universidade Federal do Rio de Janeiro (UFRJ), sendo este composto por transcritos de plântulas (3 meses após a germinação) de cana-de-açúcar do cultivar Chunnee (*S. barbari*). Assim, utilizando esses dados, pode-se buscar por Expansinas potencialmente expressas em diferentes órgãos vegetais.

Nas sequências nucleotídicas candidatas à Expansinas de cana-de-açúcar também buscou-se a presença dos domínios característicos destas proteínas. Entretanto, devido à grande parte das sequências serem apenas pedaços de transcritos, esse fator não foi determinante para a sua classificação. Os transcritos que não apresentaram ao menos um dos domínios foram confirmados através de Blast no TAIR (*The Arabidopsis Information Resource*). Com base nisso, os transcritos de cana-de-açúcar que não retornavam Expansinas como resultado do Blast em *Arabidopsis* foram excluídos das análises seguintes.

As sequências de transcritos confirmados como Expansinas foram traduzidas para proteínas por meio do programa ExPASy (www.expasy.org), sendo posteriormente editadas, retirando-se a região UTR do gene por meio de comparação com gene de sorgo ou milho mais próximo, deixando apenas a parte codificante (CDS- *Coding DNA Sequence*), uma vez que isso poderia alterar a análise filogenética posterior.

4.3 ANÁLISES FILOGENÉTICAS EM CANA-DE-AÇÚCAR

Para a construção das árvores filogenéticas com as sequências nucleotídicas de cana-de-açúcar foram utilizados genes de Expansinas de sorgo e milho como referência para a formação do arranjo da árvore. Essas sequências foram alinhadas com os transcritos de cana-de-açúcar de cada banco, pelo programa MEGA 6, utilizando a ferramenta do *Muscle*. Em seguida, a árvore foi construída pelo programa Seaview (GOUY; GUINDON; GASCUEL, 2010) empregando o pacote *PhyML* (GUINDON et al., 2010), método de Máxima Verossimilhança, utilizando os parâmetros de *bootstrap* com 1000 repetições e modelo WAG. Este modelo foi encontrado como o melhor para as análises filogenéticas por meio do *Model Test* executado pelo MEGA6.

Com base no arranjo filogenético também foi feita a denominação de grupos de sequências de cana-de-açúcar agrupadas ou próximas à um gene de sorgo ou milho, utilizando como base o código deste gene para nomear a Expansina em cana. Além disso, dentre as sequências de transcritos pertencentes a um mesmo grupo homólogo foi possível observar uma redundância entre os transcritos, sendo por isso estas sequências alinhadas e analisadas quanto à sua similaridade por meio do programa Bioedit (HALL, 1999). A partir disto, sequências em um mesmo ramo e com identidade maior que 90% foram reunidas em um único grupo (Apêndice B), onde se identificou aquela de maior tamanho para representar as demais em análises posteriores. Todos os transcritos foram identificados em uma tabela correlacionando o código da sequência nucleotídica selecionada, os transcritos que representava e a nomeação com base em sorgo ou milho. Algumas Expansinas agruparam com códigos de mesma numeração em milho e sorgo, sendo por isso adicionado letras minúsculas ao código para fins de identificação.

4.4 BUSCAS GENÔMICAS

Pesquisadores do CTBE trabalham no projeto de sequenciamento preliminar do genoma de cana a partir do cultivar SP80-3280, sendo os dados obtidos até o momento disponibilizado pelo site <http://bce.bioetanol.cnpem.br>. Desta forma, os transcritos de cana foram utilizados como sondas pela ferramenta do Blast para a identificação dos *locus* gênicos correspondentes, sendo a sequência encontrada alinhadas pelo programa Multalin (<http://multalin.toulouse.inra.fr>), onde foi possível observar a correspondência do gene com a sequência do transcrito.

Com o objetivo de verificar a presença de genes de Expansina no genoma de cana-de-açúcar que não apresentam expressão nos diferentes bancos de transcriptomas analisados, os genes de sorgo que não agruparam com nenhum transcrito em cana também foram usados para busca genômica. Assim, as sequências gênicas encontradas foram analisadas pelo programa Fgenesh (www.softberry.com), que utilizou a proteína de sorgo para encontrar a sequência codificante no genoma de cana. As sequências nucleotídicas obtidas pelo programa foram posteriormente traduzidas em proteínas pelo programa Expasy.

4.5 ANÁLISE DE EXPRESSÃO *IN SILICO*

Adicionalmente, realizou se uma análise da expressão *in silico* das sequências de Expansinas presentes no banco do SUCEST a partir de dados construídos com base na frequência de ESTs presentes nesta biblioteca.

Desta forma, foram analisados oito tecidos (Tabela 3), de diferentes bibliotecas, buscando identificar o padrão de expressão destas sequências em diferentes órgãos da planta, sendo estes: flores em diferentes estágios do desenvolvimento, folhas de plantas imaturas, folhas estioladas de plantas crescidas *in vitro*, zona radicular de plantas imaturas, raiz e ápice de plantas maduras, epiderme, primeiro e quarto entrenó de plantas imaturas e sementes em desenvolvimento.

Tabela 3- Código das diferentes amostras de tecidos vegetais analisados quanto a expressão em cana-de-açúcar.

Código Sucest	Descrição do tecido vegetal
FL1, FL3, FL4, FL5, FL8	Flores coletadas em diferentes estágios de desenvolvimento
LR1, LR2	Folhas de plantas imaturas
LV1	Folhas estioladas de plantas crescidas <i>in vitro</i>
RZ1, RZ2, RZ3	Região radicular de plantas imaturas
RT1, RT2, RT3	Raiz e ápice de plantas maduras
SB1	Epiderme
ST1, ST3	Primeiro e quarto entrenós
SD1, SD2	Sementes em desenvolvimento

Fonte: Adaptado (VETTORE et al., 2001)

Os dados de expressão foram normalizados por meio da divisão do número de leitura dos ESTs pelo total de transcritos da biblioteca e os valores anotados em logaritmo de base 2. A partir destes dados se construiu *heat maps* para cada subfamília de Expansina por meio do site HIV databases (www.hiv.lanl.gov) e fez se uso de média aritmética para encontrar um único valor de expressão para tecidos com mais de uma biblioteca. Os dados foram clusterizadas, sendo as sequências com menor expressão representadas em amarelo e com maior valor de expressão em vermelho.

4.6 ANÁLISE DA ESTRUTURA DAS SEQUÊNCIAS PROTEICAS E GENÔMICA DE EXPANSINAS DE CANA-DE-AÇÚCAR

Buscando caracterizar os domínios e os motivos conservados presentes nas proteínas Expansinas e identificar possíveis diferenças entre as subfamílias deste grupo, as sequências de EXPA, EXPB e EXPL de cana-de-açúcar e sorgo (Apêndice C) foram submetidas à análise de alinhamento múltiplo por meio do programa Clustal X (LARKIN et al., 2007). Para isso usou se como base Expansinas de tomate com a estrutura proteica (peptídeo sinal, região do domínio D1 e D2, motivos e resíduos conservados) já caracterizados pelo trabalho de Lu *et al.* (2016). Desta forma, através de uma análise comparativa, foi possível identificar motivos, resíduos conservados e a região dos domínios nas proteínas das gramíneas estudadas.

Adicionalmente, identificou se a distribuição dos motivos conservados dentro das sequências proteicas de sorgo (Apêndice E) e cana-de-açúcar, utilizando para isso o programa

MEME (<http://meme-suite.org/tools/meme>, BAILEY et al., 2009) e os parâmetros descritos nos trabalhos de Lu *et al.* (2016) e Zhu *et al.* (2014), sendo o número máximo de motivos: 10; número de repetições: *any* e tamanho: 20-100 aminoácidos.

A estrutura de éxons e íntrons foi averiguada nas sequências genômicas de Expansinas de sorgo (Apêndice F) e cana-de-açúcar. As regiões UTR de cana foram encontradas por homologia com sorgo e a localização de éxons e íntrons obtida por meio do programa Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>).

4.7 CONSTRUÇÃO DO MODELO 3D DE EXPANSINAS DE CANA-DE-AÇÚCAR

Após a realização das análises citadas anteriormente, buscou-se então construir modelos preditos da estrutura tridimensional de Expansinas de sorgo e cana-de-açúcar. Para tanto, utilizou-se como molde a estrutura de *ZmEXPB1*, já cristalizada e caracterizada experimentalmente (PDB (Protein Data Base):2HCZ) pelo trabalho de Yennawar *et al.* (2006) e as sequências de sorgo e cana-de-açúcar homólogas à este gene (*SbEXPB13* e *ShEXPB13*, respectivamente). Assim, com base em modelagem comparativa, o programa Swiss-model (swissmodel.expasy.org/) foi capaz de fornecer um modelo predito da estrutura terciária das Expansinas mencionadas, sendo estas em seguida analisadas pelo programa TopMatch (topmatch.services.came.sbg.ac.at) quanto a semelhança destas estruturas com *ZmEXPB1*. Por meio deste programa foi possível identificar um alto grau de identidade entre as estruturas tridimensionais destas espécies (<95%), o que possibilitou a comparação dos dados obtidos com o que já se tem descrito na literatura.

O modelo predito de *ShEXPA1* foi obtido com base na estrutura de *ZmEXPB1* (PDB: 2HCZ), empregando uma metodologia similar à usada por Gaete-Eastman *et al.* (2015) em seu trabalho com proteínas EXPA e EXPB de espécie *Vasconcellea pubescens*, uma eudicotiledônea.

4.8 COLETA DO MATERIAL VEGETAL

Para a realização deste trabalho foram coletadas amostras de híbridos de cana-de-açúcar com teores contrastantes de hemicelulose, celulose e lignina, sendo estes já previamente caracterizadas pelo estudo de Masarin *et. al* (2011).

Os híbridos de cana-de-açúcar foram plantados no Campus 1 da Escola de Engenharia de Lorena (EEL-USP) e cultivadas em ambiente aberto por aproximadamente um ano, até atingirem a maturidade. Dentre os estes híbridos, os números 140 e 89 foram escolhidos para as análises devido às características bioquímicas contrastantes entre eles, sendo este último muito usado em estudos de aumento de biomassa e digestibilidade (LAURITO-FRIEND *et al.*, 2015). Após o período de crescimento das plantas foi feita a coleta de amostras de diferentes estágios de desenvolvimento do colmo. Para isso, a coleta foi feita com base no artigo de Zhang *et al.* (2014a) realizando a contagem a partir do primeiro entrenó da planta, numerando os demais de cima para baixo, sendo os primeiros os mais jovens e os últimos mais maduros. Para as análises posteriores buscou-se amostras de diferentes estágios de maturação, sendo por isso coletados os entrenós 1, 3, 5, 7 e 9, de onde retirou-se a epiderme e separou o córtex e a medula do colmo. Esses dois tecidos foram congelados em nitrogênio líquido, sendo em seguida armazenados em freezer -80°C. Este procedimento foi feito em triplicata para os dois híbridos coletados.

Adicionalmente amostras de raiz, caule e folha de plantas jovens (2 meses após o plantio) dos híbridos 140 e 89, crescidas em vermiculita, foram coletadas, congeladas em nitrogênio líquido e mantidas em freezer -80°C.

4.9 DESENHO DOS INICIADORES E TESTE DE ESPECIFICIDADE POR PCR

Para as futuras análises de expressão por PCR em tempo real dos transcritos de Expansinas em cana-de-açúcar realizou-se o desenho dos iniciadores selecionando entre o grupo de sequências homólogas de um mesmo ramo filogenético aquelas com maior tamanho. A construção dos iniciadores foi feita empregando o programa Primer 3 (bioinfo.ut.ee/primer3-0.4.0/), obtendo-se o *primer forward* e o *primer reverse* (Apêndice G).

Por meio do programa Primer 3 também pode se desenhar os iniciadores na região 3' UTR, o que é vantajoso já que essa parte da sequência é diferente para cada gene da família, aumentando dessa forma a especificidade do par de iniciadores. Assim, as sequências nucleotídicas de Expansinas de cana-de-açúcar que possuíam essa parte do gene tiveram seus *primers* desenhados nessa região.

Após o desenho dos iniciadores se prosseguiu para a etapa de maceração das amostras já previamente coletadas dos entrenós e medulas 1, 3, 5, 7 e 9 dos híbridos 89 e 140, assim como as amostras de folha e raiz do híbrido 140. A ruptura da parede celular foi feita através do macerador *Geno/grinder* à baixa temperatura utilizando nitrogênio líquido. Posteriormente foi utilizado cerca de 300 mg das amostras maceradas para a extração de RNA total, sendo para isso empregado o reagente *Concert™ Plant RNA Reagent (Life Technologies TM)*, seguindo as instruções fornecidas pelo fabricante. O RNA extraído teve a sua concentração e relações de absorbância 260nm/280nm e 260nm/230nm quantificadas através do espectrofotômetro *Biodrop* e sua integridade avaliada através de eletroforese em gel de agarose 1%. As amostras de RNA que apresentaram boas relações de absorbância e sua integridade confirmada foram tratadas com DNase e convertidas em cDNA por meio do kit *SuperScript^R III Reverse Transcriptase (Thermo Fisher Scientific)* segundo as especificações do fabricante.

Buscando avaliar a especificidade dos iniciadores das sequências de Expansinas em cana-de-açúcar antes de realizar análises de PCR em Tempo Real, primeiramente foi realizado a validação dos *primers* da subfamília EXPB. Para isso se empregou a técnica de PCR convencional em amostras de medula do entrenó 5 e 3 dos híbridos 140.1 e 140.2, respectivamente, já previamente extraídas e convertidas em cDNA pelo mesmo método citado anteriormente. A realização da análise de PCR foi feita por meio do kit *Taq DNA Polimerase, recombinant (Thermo Fisher Scientific)*, seguindo o protocolo fornecido pelo fabricante, sendo em seguida o resultado da amplificação avaliado por eletroforese em gel de agarose 3%.

5 RESULTADOS E DISCUSSÃO

5.1 IDENTIFICAÇÃO DE EXPANSINAS EM ESPÉCIES RELACIONADAS À CANA-DE-AÇÚCAR

Por meio da ferramenta do BLASTp disponível no banco de dados do PHYTOZOME, foi possível encontrar diversos genes candidatos à Expansinas no genoma de espécies relacionadas a cana-de-açúcar: *Zea mays*, *Sorghum bicolor* e *Brachypodium distachyon*. Após a confirmação da presença dos domínios característicos das Expansinas vegetais, DPBB e *Pollen Allergen*, realizou-se a construção das árvores filogenéticas. As análises revelaram a presença de clados separados e visivelmente distintos, tornando possível a separação das subfamílias gênicas de EXPA, EXPB e EXPL nas diferentes espécies de plantas (Figuras 6 a 8). A construção da árvore filogenética foi realizada tanto com as sequências nucleotídicas quanto as de proteínas, no entanto, apresentou-se neste trabalho a árvore obtida por meio das sequências proteicas, já que estas apresentam menor variação dentro de sua cadeia constituinte.

Figura 6- Árvore filogenética das sequências proteicas de Expansinas encontradas em *Zea mays* sp. Destacado em vermelho estão os ramos com os genes de EXPA, em azul EXPB e verde EXPL.

Fonte: Arquivo pessoal

Figura 7- Árvore filogenética das sequências proteicas de Expansinas encontradas em *Sorghum bicolor* sp. Destacado em vermelho estão os ramos com os genes de EXPA, em azul EXPB e verde EXPL.

Fonte: Arquivo pessoal

Figura 8- Árvore filogenética das sequências proteicas de Expansinas encontradas em *Brachypodium distachyon* sp. Destacado em vermelho estão os ramos com os genes de EXPA, em azul EXPB e verde EXPL.

Fonte: Arquivo pessoal

Através da comparação entre os genes de Expansinas encontradas nesta análise com aquelas já descritas em milho por Zhang *et al.* (2014) foi possível realizar a validação do método empregado para este estudo, uma vez que resultados semelhantes foram obtidos. Na subfamília EXPA houve uma pequena diferença entre os resultados, uma vez que foram identificados trinta e seis genes no artigo de Zhang *et al.* (2014) e trinta e nove no presente trabalho. Esta pequena variação também ocorreu nas subfamílias EXPB e EXPL, onde obteve-se um gene a menos em EXPB (total de 47 genes), e um a mais em EXPL (total de 5 genes) (Tabela 4).

Um dos motivos para a variação entre esses dois estudos pode ser à diferença entre os bancos utilizados, uma vez que o trabalho de Zhang *et al.* (2014) utilizou o banco disponível no site <http://ftp.maizesequence.org/current>, enquanto as análises do presente projeto foram realizadas pelo banco do PHYTOZOME. Já o gene a menos encontrado em EXPB deve-se ao fato do uso de estratégias diferentes, uma vez que Zhang *et al.* (2014) utilizam sequências de isoformas, ou seja, variações de um mesmo *locus* gênico, que possuem pequenas mudanças de aminoácidos, enquanto as sequências proteicas consideradas nesta análise foram apenas uma única isoforma de cada gene.

Já as análises globais no genoma buscando a caracterização da superfamília gênica de Expansinas nas espécies *Sorghum bicolor* e *Brachypodium distachyon* foram realizadas primeiramente neste estudo. Os trabalhos publicados até o momento relacionados à caracterização de Expansinas em sorgo focaram na identificação de genes específicos, expressos em determinados órgãos em condições de estresse, como é o caso de Buchanan *et al.* (2005), que observaram transcritos de EXPB4 em resposta à desidratação, alta salinidade e a presença de ácido abscísico (ABA) na raiz. Já em *Brachypodium distachyon*, um estudo global do proteoma desta espécie identificou a presença de Expansinas, porém, não de forma a caracterizar os genes desta família (WANG *et al.*, 2010). Dessa forma, a identificação de genes de Expansinas no genoma de diferentes espécies de gramíneas realizada neste trabalho permitirá a ampliação dos estudos desta família gênica.

Assim, com base nas análises realizadas, a quantidade de genes de EXPA, EXPB e EXPL encontradas nas três espécies analisadas foi representada na Tabela 4.

Tabela 4- Quantidade de genes de Expansinas identificadas em cada subfamília gênica em diferentes espécies de gramíneas

Colunas1	Arroz*	Milho	Sorgo	<i>Brachypodium</i>
EXPA	34	39	36	32
XPB	19	47	37	24
EXPL	5	5	4	6
Total	58	91	77	62

*Sequências retiradas do site www.personal.psu.edu/fsl/ExpCentral

Fonte: Arquivo pessoal

Comparando esses dados com os encontrados no genoma de plantas eudicotiledôneas é possível identificar a expansão da subfamília EXPB em gramíneas. Em *Arabidopsis*, por exemplo, são encontrados 6 genes EXPB em contraste com 26 genes de EXPA. O mesmo pode ser observado em soja e *Populus* que possuem 49 e 27 genes EXPA e apenas 9 e 3 genes em EXPB, respectivamente (SAMPEDRO et al., 2015; ZHU et al., 2014a). De acordo com Sampedro et al. (2015) o aumento do número de genes da subfamília EXPB em gramíneas deve se a diversificação do clado EXPB-I, o que é evidenciado pela radiação de 16 genes de β -expansinas em arroz comparado com apenas 4 em *Arabidopsis* e um em *Populus*.

Após a identificação das Expansinas pertencentes as três subfamílias desse grupo, também foram realizadas análises filogenéticas pelo MEGA6 englobando as quatro espécies de gramíneas estudadas: *Oryza sativa*, *Zea mays*, *Sorghum bicolor* e *Brachypodium distachyon* (Figuras 9 a 11). Com base no arranjo observado na árvore filogenética de EXPB é possível identificar os clados citados por Sampedro et al. (2015), evidenciando os genes do clado EXPB-I conservado das monocotiledôneas (Figura 10) e dois ramos distintos identificados como divergentes, exclusivos de gramíneas e espécies próximas. O clado nomeado EXPB-I divergente vegetativo está relacionado à Expansinas expressas em tecidos vegetativos, enquanto o clado EXPB-I divergente de pólen está relacionado com Expansinas preferencialmente expressas em grão de pólen (Figura 10).

Figura 9- Árvore filogenética com as sequências proteicas de EXPA encontradas em arroz (verde), milho (azul), sorgo (vermelho), *Brachypodium distachyon* (amarelo) e *Arabidopsis thaliana* (roxo).

Fonte: Arquivo pessoal

Figura 10- Árvore filogenética com as sequências proteicas de EXPB encontradas em arroz (círculos verdes), milho (círculos azuis), sorgo (círculos vermelhos), *Brachypodium distachyon* (círculos amarelos) e *Arabidopsis thaliana* (roxo).

Fonte: Arquivo pessoal

Figura 11- Árvore filogenética com as sequencias proteicas de EXPL encontradas em arroz (verde), milho (azul), sorgo (vermelho), *Brachypodium distachyon* (amarelo) e *Arabidopsis thaliana* (roxo).

Fonte: Arquivo pessoal

5.2 IDENTIFICAÇÃO DE EXPANSINAS EM CANA-DE-AÇÚCAR

Após a busca de Expansinas em espécies de gramíneas próximas à cana-de-açúcar, as sequências gênicas encontradas em sorgo foram usadas como sondas, empregando a ferramenta do Blast em sete bancos de transcriptomas de cana-de-açúcar. Os transcritos obtidos foram analisados e confirmados como pertencentes a família das Expansinas por ferramentas de bioinformática. Desta forma, chegou se ao número total de 190 transcritos obtidos no presente estudo, sendo 36 sequências encontradas no banco de Nishiyama *et al.* (2014); 39 no SUCEST; 29 na biblioteca de ESTs e RNA-seq de Graviton *et al.* (2014), 11 transcritos no conjunto de RNAseq de Cardoso-Silva *et al.* (2014), 22 no RNA seq cedido pela professora Adriana Hemerly (UFRJ); 26 sequências nucleotídicas no banco de RNA seq disponibilizado pelo CTBE (Laboratório Nacional de Ciência e Tecnologia do Bioetanol) e 27 no banco de transcritos expressos no entrenó (VICENTINI *et al.*, 2015)(Figura 12).

Na figura 12 também está representado a quantidade de transcritos encontrados nas três subfamílias gênicas de Expansinas nas diferentes bibliotecas analisadas. Com base nesta figura também pode-se notar que, de forma geral, o número de sequências nucleotídicas seguem o padrão esperado para as gramíneas, sendo perceptível o aumento do número de transcritos EXPB, que se aproximam ou ultrapassam a quantidade de transcritos EXPA, o que pode estar relacionado a maior especificidade destas proteínas sobre a parede celular das gramíneas (COSGROVE, 2000b; WU; MEELEY; COSGROVE, 2001).

Figura 12- Transcritos encontrados nos diferentes bancos analisados. 1- Folhas, entrenós imaturos e intermediários da cv. SP80-;3280 (Nishiyama, *et al*, 2014); 2- SUCEST (VETTORE *et al.*, 2001); 3- Folhas jovens da cv. SP70-1143 (Grativol, *et al*, 2014); 4- Folhas de 6 cv. (Cardoso-Silva, *et al*, 2014) ; 5- Plântula (Cedido pela prof. Adriana Hemerly-UFRJ); 6- banco de transcritos de diferentes segmentos de folhas da cultivar SP80-3280 (CTBE); 7- banco de transcritos expressos no entrenó dos cultivares IACSP04-065 e IACSP04-627 (VICENTINI *et al.*, 2015)

Fonte: Arquivo pessoal

Com o término da procura pelos transcritos, foi verificada uma provável redundância entre os diferentes bancos, sendo por isso analisando detalhadamente os transcritos que se agruparam em um mesmo ramo homólogo da árvore filogenética. Por meio de alinhamento dessas sequências com o gene de sorgo ou milho mais próximo filogeneticamente foi possível observar grande similaridade entre elas, além de transcritos que se encontravam em pequenos pedaços, mas que apresentam sobreposição originando os chamados *contigs*.

Foi então realizado um teste de identidade destas sequências por meio do programa Bioedit (HALL, 1999), onde sequências com mais de 90% de similaridade foram consideradas membros de um mesmo grupo homólogo (Apêndice B). Baseado nisso, foi possível escolher um transcrito de maior tamanho que pudesse representar as demais sequências nucleotídicas relacionadas a este, para assim facilitar os trabalhos posteriores e o desenho dos iniciadores. Após esta organização, os transcritos de cana selecionados foram nomeados com base no gene de sorgo ou milho mais próximo filogeneticamente. Adicionalmente, foi realizado a análise no banco parcial de dados genômicos de cana-de-açúcar, disponibilizados pelo CTBE. Desta forma, os transcritos foram utilizados como sondas pela ferramenta do Blast, visando a identificação dos *locus* gênicos correspondentes.

Com o objetivo de verificar a presença de genes de Expansina no genoma de cana-de-açúcar que não apresentam expressão nos diferentes transcriptomas analisados, os genes de sorgo que não agruparam com nenhum transcrito em cana também foram usados para busca no genoma. No entanto, dentre as sequências de sorgo utilizadas, algumas não encontraram correspondente homólogo no genoma disponível, ocasionando lacunas na nomeação das sequências nucleotídicas obtidas em cana em relação aos genes de sorgo.

Os dados obtidos para EXPA, EXPB e EXPL estão organizados nas tabelas de 5 a 7, respectivamente, indicando o nome dado ao maior transcrito de Expansina de cana-de-açúcar encontrado em um grupo filogenético, seu homólogo em sorgo ou milho, o código do fragmento genômico onde está localizado o *locus* do gene e o número de transcritos encontrados nos diferentes bancos analisados.

Tabela 5- Nomeação dos transcritos EXPA em cana-de-açúcar, correlacionados com o homólogo no genoma de sorgo e milho, o código do fragmento do genoma correspondente e o número de transcritos encontrados nos 7 bancos analisados.

Código em cana	Homólogo	Código do gene	Número de transcritos
<i>ShEXPA1-sb</i>	<i>SbEXPA1</i>	lcl SCSP803280_000003966	7
<i>ShEXPA1-zm</i>	<i>ZmEXPA1</i>	-	18
<i>ShEXPA4</i>	<i>SbEXPA4</i>	lcl SCSP803280_000000780	-
<i>ShEXPA5</i>	<i>SbEXPA5</i>	lcl SCSP803280_000096313	2
<i>ShEXPA8</i>	<i>SbEXPA8</i>	lcl SCSP803280_000039302	1
<i>ShEXPA9</i>	<i>SbEXPA9</i>	lcl SCSP803280_000017917	-
<i>ShEXPA12</i>	<i>SbEXPA12</i>	-	1
<i>ShEXPA13</i>	<i>SbEXPA13</i>	lcl SCSP803280_000190559	1
<i>ShEXPA14</i>	<i>SbEXPA14</i>	lcl SCSP803280_000043058	1
<i>ShEXPA15</i>	<i>ZmEXPA15</i>	-	4
<i>ShEXPA16</i>	<i>SbEXPA16</i>	lcl SCSP803280_000033064	3
<i>ShEXPA17</i>	<i>SbEXPA17</i>	lcl SCSP803280_000012816	4
<i>ShEXPA19</i>	<i>SbEXPA19</i>	lcl SCSP803280_000111101	4
<i>ShEXPA24</i>	<i>ZmEXPA24</i>	-	1
<i>ShEXPA25</i>	<i>SbEXPA25</i>	-	2
<i>ShEXPA27-sb</i>	<i>SbEXPA27</i>	lcl SCSP803280_000034776	6
<i>ShEXPA27-zm</i>	<i>ZmEXPA27</i>	-	2
<i>ShEXPA28</i>	<i>SbEXPA28</i>	lcl SCSP803280_000042948	-
<i>ShEXPA31</i>	<i>SbEXPA31</i>	lcl SCSP803280_000101887	1
<i>ShEXPA32</i>	<i>SbEXPA32</i>	lcl SCSP803280_000006787	-
<i>ShEXPA33</i>	<i>SbEXPA33</i>	-	3
<i>ShEXPA34</i>	<i>SbEXPA34</i>	lcl SCSP803280_000001972	-
<i>ShEXPA35</i>	<i>SbEXPA35</i>	lcl SCSP803280_000171415	3
<i>ShEXPA36</i>	<i>SbEXPA36</i>	lcl SCSP803280_000095130	1

Fonte: Arquivo pessoal

Tabela 6- Nomeação dos transcritos EXPB em cana-de-açúcar, correlacionados com o homólogo no genoma de sorgo e milho, o código do fragmento do genoma correspondente e o número de transcritos encontrados nos 7 bancos analisados.

Código em cana	Homólogo	Código do gene	Número de transcritos
<i>ShEXPB1</i>	<i>SbEXPB1</i>	lcl SCSP803280_000145538	6
<i>ShEXPB2</i>	<i>SbEXPB2</i>	-	4
<i>ShEXPB3-zm</i>	<i>ZmEXPB3-α</i>	-	4
<i>ShEXPB3-sb</i>	<i>SbEXPB3</i>	-	3
<i>ShEXPB4</i>	<i>SbEXPB4</i>	lcl SCSP803280_000189163	5
<i>ShEXPB5</i>	<i>SbEXPB5</i>	-	2
<i>ShEXPB6</i>	<i>SbEXPB4</i>	-	3
<i>ShEXPB9</i>	<i>SbEXPB9</i>	lcl SCSP803280_000155728	2
<i>ShEXPB10</i>	<i>SbEXPB10</i>	-	1
<i>ShEXPB11</i>	<i>SbEXPB11</i>	lcl SCSP803280_000076988	-
<i>ShEXPB12</i>	<i>SbEXPB12</i>	-	2
<i>ShEXPB13</i>	<i>SbEXPB13</i>	lcl SCSP803280_000023980	-
<i>ShEXPB17</i>	<i>SbEXPB17</i>	lcl SCSP803280_000043348	-
<i>ShEXPB18</i>	<i>SbEXPB18</i>	lcl SCSP803280_000072106	-
<i>ShEXPB19</i>	<i>SbEXPB19</i>	-	1
<i>ShEXPB21</i>	<i>SbEXPB21</i>	-	5
<i>ShEXPB22</i>	<i>SbEXPB22</i>	lcl SCSP803280_000113424	1
<i>ShEXPB23</i>	<i>SbEXPB23</i>	lcl SCSP803280_000108437	12
<i>ShEXPB24</i>	<i>SbEXPB24</i>	lcl SCSP803280_000113567	1
<i>ShEXPB25</i>	<i>SbEXPB25</i>	lcl SCSP803280_000040604	7
<i>ShEXPB26</i>	<i>SbEXPB26</i>	-	4
<i>ShEXPB27</i>	<i>SbEXPB27</i>	-	1
<i>ShEXPB28</i>	<i>SbEXPB28</i>	lcl SCSP803280_000065540	6
<i>ShEXPB29</i>	<i>SbEXPB29</i>	lcl SCSP803280_000147924	1
<i>ShEXPB30</i>	<i>SbEXPB30</i>	lcl SCSP803280_000088941	-
<i>ShEXPB32</i>	<i>SbEXPB31</i>	lcl SCSP803280_000019559	3
<i>ShEXPB33</i>	<i>SbEXPB33</i>	lcl SCSP803280_000064994	2

Fonte: Arquivo pessoal

Tabela 7- Nomeação dos transcritos EXPL em cana-de-açúcar, correlacionados com o homólogo no genoma de sorgo e milho, o código do fragmento do genoma correspondente e o número de transcritos encontrados nos 7 bancos analisados.

Código em cana	Homólogo	Código do gene	Número de transcritos
<i>ShEXPL1</i>	<i>SbEXPL1</i>	-	4
<i>ShEXPL2</i>	<i>SbEXPL2</i>	lcl SCSP803280_000124172	6
<i>ShEXPL3</i>	<i>SbEXPL3</i>	lcl SCSP803280_000095044	8
<i>ShEXPL4</i>	<i>SbEXPL4</i>	lcl SCSP803280_000037034	5

Fonte: Arquivo pessoal

Com base nas tabelas apresentadas pode-se constatar que para a maioria das sequências de transcritos de Expansinas encontradas em cana-de-açúcar foi identificado o gene correspondente. Os genes que não foram encontrados, mas possuem transcritos podem ainda não terem sido sequenciados pelo grupo de pesquisa do CTBE, uma vez que o sequenciamento do genoma ainda não está completo.

No total foram identificadas 24 sequências nucleotídicas de Expansina nas subfamílias EXPA, 27 em EXPB e 4 em EXPL em cana-de-açúcar. Este número mostra-se inferior ao esperado quando comparado à EXPA e EXPB de milho e sorgo, porém similares em EXPL (Tabela 4). Isto sugere que ainda não foram identificadas todas as sequências nucleotídicas da família das Expansinas em cana, entretanto estes dados já fornecem uma base para novos estudos nesta área.

Posteriormente, as sequências nucleotídicas de Expansinas em cana-de-açúcar foram traduzidas para proteína e com base nas cadeias peptídicas foram construídas as árvores filogenéticas juntamente com as sequências proteicas de sorgo, milho, arroz e *Arabidopsis* (Figuras 13 a 15). Por meio das análises filogenéticas é possível observar a relação de ortologia entre os genes das diferentes espécies. Devido aos genes ortólogos provavelmente desempenharem funções equivalentes entre os organismos, sua identificação é uma importante ferramenta no estudo funcional de genes (KOONIN, 2005). Desta forma, através do arranjo formado na árvore filogenética foi possível distinguir as relações de ortologia entre as espécies, permitindo a identificação de sequências nucleotídicas com possíveis funções equivalentes (KOONIN, 2005).

No trabalho de Seki *et al.* (2014) as proteínas *OsEXPA33*, *OsEXPB8* e *OsEXPL3* demonstraram maior efeito no aumento da degradação da celulose cristalina. Assim, baseado

na formação dos ramos filogenéticos destes genes com cana-de-açúcar, destacado pelos quadrados vermelhos nas figuras 13, 14 e 15, as Expansinas *ShEXPA14*, *ShEXPA24*, *ShEXPB22* e *ShEXPL3* podem apresentar atividade similar no aumento da sacarificação.

Figura 13: Árvore filogenética da subfamília EXPA com sequências proteicas de cana-de-açúcar (vermelho), sorgo (azul), milho (verde), arroz (roxo) e *A. thaliana* (azul claro). Destacado pelo quadrado vermelho estão os ramos com sequências de cana homólogas às sequências de arroz com interesse biotecnológico.

Fonte: Arquivo pessoal

Figura 14: Árvore filogenética da subfamília EXPB com sequências proteicas de cana-de-açúcar (vermelho), sorgo (azul escuro), milho (verde), arroz (roxo) e *A. thaliana* (azul claro). Destacado pelo quadrado vermelho estão os ramos com sequências de cana homólogas às sequências com interesse biotecnológico.

Figura 15- Árvore filogenética da família EXPL com sequências proteicas de cana-de-açúcar (vermelho), sorgo (azul), milho(verde); arroz (roxo) e *A. thaliana* (azul claro). Destacado pelo quadrado vermelho estão os ramos com sequências de cana homólogas às sequências de arroz com interesse biotecnológico.

Fonte: Arquivo pessoal

No estudo de Ma *et al.* (2013) foi demonstrado que a superexpressão da Expansina *OsEXPA8* levou ao aumento do tamanho da planta de arroz e também à uma maior massa radicular. Desta forma, como as Expansinas de cana *ShEXPA24* e *ShEXPA14* são mais próximas filogeneticamente deste gene de arroz estas sequências também podem apresentar uma atividade de aumento de biomassa, sendo portanto bons candidatos a serem usados em estudos com foco nos experimentos de aumento de biomassa em cana-de-açúcar.

No trabalho de Buchanan *et al.* (2005) também foi identificada a transcrição do gene *SbEXPB4* em sorgo, homólogo a *ShEXPB4*, durante a resposta vegetal à desidratação, alta salinidade e presença de ácido abscísico. Desta forma, a sequência nucleotídica de cana *ShEXPB4* também pode apresentar grande relevância em estudos com foco na resistência de cana-de-açúcar ao estresse hídrico.

Estes dados permitem a seleção de sequências nucleotídicas de interesse que podem levar tanto a uma melhora do processo de sacarificação para a produção de biocombustíveis, como também ao aumento da biomassa da cana-de-açúcar.

5.3 ANÁLISE DA SEQUÊNCIA, DOMÍNIOS E MOTIVOS CONSERVADOS DAS EXPANSINAS

Buscando caracterizar a estrutura dos genes de cana-de-açúcar pertencentes às subfamílias EXPA, EXPB e EXPL foi feita a análise da estrutura de éxons e íntrons dos genes de sorgo (Apêndice F) e cana-de-açúcar (Figura 16, 17 e 18). Nesta análise, devido à ausência de anotação, as regiões UTR's dos genes de cana foram encontradas pela identidade com seus homólogos em sorgo. Apesar da maioria dos genes de cana apresentarem uma região UTR com tamanho aproximado ao anotado em sorgo, alguns genes apresentaram alterações no tamanho da região UTR em relação ao seu homólogo (5 em EXPA, 2 em EXPB e 1 em EXPL), sendo alguns destes devido à anotação parcial do genoma de cana.

A partir dos resultados preliminares obtidos é possível observar que, de forma geral, sequências dentro de uma mesma subfamília tem uma estrutura de éxons e íntrons parecida, porém diferentes padrões são encontrados entre EXPA, EXPB e EXPL, sendo este dado também observado no trabalho de Zhu *et al.* (2014) em sua análise da estrutura de Expansinas em soja. Em EXPA de cana-de-açúcar a maioria das sequências (10 de 17) possuem estrutura de 2 éxons, enquanto 3 possuem 3 éxons e 4 possuem apenas 1 éxon. Um padrão similar à esse foi encontrado também em sorgo (Apêndice F) onde 20 dos 36 genes de EXPA possuem 2 éxons; 13 possuem 3 éxons; 2 possuem 4 éxons e apenas 1 possui 1 éxon em sua estrutura.

Figura 16- Análise de éxons e íntrons em EXPA de cana-de-açúcar.

Fonte: Arquivo pessoal

Em EXPB, a maioria das sequências apresentou um padrão estrutural de 4 éxons e 3 introns tanto em sorgo quanto em cana (22 de 36 e 8 de 16, respectivamente), no entanto outras conformações também foram encontradas, sendo 4 genes em sorgo e 4 em cana com 3 éxons, 4 genes em sorgo e 2 em cana com 2 éxons e por fim 6 genes de sorgo e 1 de cana com apenas um éxon. Padrões similares aos encontrados neste estudo também foram observados na estrutura de soja e tomate, onde se encontrou um perfil majoritário de 3 éxons nas sequências de EXPA e 4 éxons em todas as EXPBs (LU et al., 2016; ZHU et al., 2014a).

Estas variações dentro das subfamílias das Expansinas, tanto em EXPA e EXPB podem ser o resultado de ganho ou perda de introns ao longo da evolução. Adicionalmente, o alto grau de similaridade estrutural e identidade entre os genes de uma mesma subfamília também pode estar relacionada aos eventos de duplicação dentro do genoma das plantas (LU et al., 2016; ZHU et al., 2014a).

Figura 17- Análise de éxons e introns em EXPB de cana-de-açúcar.

Fonte: Arquivo pessoal

Já em EXPL, os genes de sorgo apresentaram dois padrões gerais, sendo 2 sequências com 5 exons e 2 com 4 éxons. Em cana-de-açúcar no entanto, houve uma maior variação de estrutura, não sendo possível encontrar um padrão definido. Estes dados são condizentes com o trabalho de Zhu *et al.* (2014a) que observou que as sequências de EXPL variam mais quanto à sua estrutura do que EXPA e EXPB, mostrando uma maior conservação destas subfamílias possivelmente devido a maior pressões seletiva, em comparação com o grupo das expansinas-like.

Figura 18- Análise de éxons e introns em EXPL de cana-de-açúcar.

Fonte: Arquivo pessoal

De forma à caracterizar de maneira mais detalhada as subfamílias de Expansinas em cana e sorgo foi realizada uma análise de domínios e motivos conservados nas sequências proteicas, uma vez que esta é uma importante ferramenta na identificação funcional de regiões das proteínas (LU et al., 2016). Desta forma, foi realizado um alinhamento das

sequências de EXPA destas espécies, assim como das sequências de EXPB, juntamente com as de tomate já previamente caracterizadas pelo trabalho de Lu *et al.*(2016).

O alinhamento múltiplo realizado pelo programa Clustal X (LARKIN et al., 2007) entre as sequências proteicas de cana e tomate (Figura 19 e 21), assim como o alinhamento das sequências de sorgo e tomate (Apêndice C), permitiu uma comparação dos dados com o trabalho de Lu *et al.*(2016) e a partir disto identificar a localização do peptídeo sinal, o domínio D1 e por fim o domínio D2 das EXPAs e EXPBs de sorgo e cana-de-açúcar. Com cerca de 20-30 aminoácidos o peptídeo sinal inicia-se na extremidade N-terminal e prolonga-se até um resíduo conservado de triptófano (W). A partir deste resíduo há o início do domínio D1 que estende-se até um resíduo conservado de cisteína (C), seguido então de uma sequência de aminoácidos que funcionam como um *linker* até um resíduo conservado de fenilalanina (F) que demarca o início da região do domínio D2 até o fim da proteína (Figuras 19 e 21).

Após a localização dos dois domínios característicos das Expansinas dentro da estrutura primária das proteínas de sorgo e cana também foi possível identificar motivos já descritos em outros trabalhos como altamente conservados nas subfamílias EXPA e EXPB, sendo estes: HFD, GGACG e TWYG (Figuras 19 e 20). Supõe-se que estes motivos estariam relacionados à formação de um sulco de ligação à polissacarídeos, porém dentre estes, a sequência composta de histidina, fenilalanina e ácido aspártico (HFD) é a mais bem caracterizada por ser parte do sítio ativo encontrado nas GH45, porém sem função catalítica nas Expansinas. Este motivo é altamente conservado tanto em monocotiledôneas quanto em eudicotiledôneas, podendo indicar essa região como crucial para a atividade das Expansinas de plantas (LU et al., 2016; YENNAWAR et al., 2006). No entanto, apesar do motivo HFD ser detectado em grande parte das EXPBs e em todas as EXPAs de soja e tomate, há a perda deste motivo na subfamília EXPL (LU et al., 2016; ZHU et al., 2014a), dado também observado em sorgo (Apêndice C) e cana-de-açúcar, o que sugere um modo de ação diferente de EXPL comparado com as demais subfamílias.

Uma característica interessante foi observada no motivo TWYG (Treonina, Triptófano, Tirosina e Glicina), que apesar de ser indicado como conservado de maneira geral nas Expansinas (LIU et al., 2007; YENNAWAR et al., 2006), mostra-se presente apenas nas sequências de EXPB, enquanto que em EXPA ocorre a substituição de um dos aminoácidos, formando um motivo conservado de TFYG (Figura 19). Esta diferença entre EXPA e EXPB pode acarretar em pequenas alterações na estrutura da proteína, que contribuem para à maior

especificidade de β -expansinas à parede celular das monocotiledôneas e de α -expansinas à eudicotiledôneas (SAMPEDRO et al., 2015).

Além destes motivos, também pôde-se distinguir resíduos conservados de cisteína e triptófano ao longo da sequência (Figuras 19 a 21) já reportados em outros estudos com eudicotiledôneas e que mostram-se presentes também em monocotiledôneas. Estes resíduos foram descritos como envolvidos na formação das pontes dissulfeto e contribuem para a conformação tridimensional destas proteínas (LU et al., 2016; SAMPEDRO; COSGROVE, 2005b; YENNAWAR et al., 2006).

Adicionalmente, diversos trabalhos identificam inserções específicas nas subfamílias EXPA e EXPB, chamadas de α e β -inserção, respectivamente (LU et al., 2016; SAMPEDRO; COSGROVE, 2005b, LI et al., 2002). Em comparação com estes dados também foi possível identificar neste trabalho a α -inserção e β -inserção em sorgo (Apêndice C) e cana-de-açúcar. A α -inserção está presente apenas na subfamília EXPA (Figura 19), já a β -inserção está presente tanto em EXPB quanto em EXPL e contém um resíduo conservado de glicina (G) em seu final (LU et al., 2016).

Figura 19- Alinhamento múltiplo das sequências de EXPA de cana-de-açúcar (*ShEXPA*) com sequências de EXPB (*SlEXPB*) e EXPA (*SlEXPA*) de tomate. Letras em vermelho destacam a localização dos motivos e resíduos conservados das sequências; setas indicam localização dos domínios e linhas representam a região do peptídeo sinal e as α e β -inserções.

Fonte: Arquivo pessoal

Figura 20- Alinhamento múltiplo das sequências de EXPB em cana-de-açúcar (*Sh*EXPB) com sequências de EXPA (*S*/EXPA) e EXPB (*S*/EXPB) de tomate. Letras em vermelho destacam a localização dos motivos e resíduos conservados das sequências; setas indicam localização dos domínios e linhas representam a região do peptídeo sinal e as α e β -inserções

Fonte: Arquivo pessoal

Estudos indicam que as sequências da subfamília EXPL possuem um padrão de estrutura que difere em alguns aspectos das demais subfamílias, sendo por esse motivo realizado um alinhamento apenas com sequências de EXPL de cana-de-açúcar e tomate (Figura 21). Utilizando como base o trabalho de Lu *et al.* (2016) foi possível identificar os domínios D1 e D2 característico das Expansinas de plantas na subfamília EXPL, como indicado na figura 21, porém devido aos diferentes padrões das sequências não foi possível delimitar a região do peptídeo sinal.

Figura 21- Alinhamento múltiplo das sequências de EXPL e cana-de-açúcar (*ShEXPL*) com sequências EXPL (*SlEXPL*) de tomate. Setas indicam as regiões dos domínios D1 e D2.

Fonte: Arquivo pessoal

Além da análise de alinhamento múltiplo também foi obtido o perfil de motivos presentes nas sequências proteicas, através do programa MEME (BAILEY *et al.*, 2006). Assim, a análise permitiu identificar o tipo, a ordem e o número de motivos conservados das subfamílias de Expansinas de sorgo (Apêndice E) e cana-de-açúcar (Figuras 22, 23 e 24), porém sem a associação destas sequências com uma função específica. Utilizando os parâmetros empregados no trabalho de Zhu *et al.* (2014a) buscou-se os 10 motivos conservados em todas as três subfamílias gênicas das Expansinas (Apêndice D) e, a partir do resultado obtido, observou-se uma distribuição mais constante dos motivos em EXPA do que EXPB e EXPL (Figuras 22, 23 e 24). Apesar de algumas sequências de cana-de-açúcar estarem incompletas foi possível encontrar a mesma ordem de motivos obtidos em sequências proteicas completa.

Em EXPB observou-se a substituição de alguns motivos, como é o caso do motivo 6, que é substituído pelo 10 em algumas β -expansinas de cana, enquanto em EXPL ocorre tanto a substituição quanto a inserção de motivos, sugerindo desta forma uma maior variação entre as sequências desta subfamília em comparação com EXPA (Figuras 23 e 24). Este fato também é observado em soja, onde se constatou que a maior parte das sequências de EXPA

compartilhavam um mesmo motivo, enquanto houve perda de dois motivos em EXPB e EXPL (ZHU *et al.*2014).

Assim, de forma geral, os motivos seguem padrões próprios em cada subfamília. Estes resultados são corroborados pelo trabalho de Zhu *et al.*(2014), que verificou em sua análise que o tipo, a ordem e o número de motivos são similares em proteínas de uma mesma subfamília, mas diferente em sequências proteicas de outras subfamílias. Estes autores descrevem que a distribuição dos motivos das EXPA em soja é significativamente diferente se comparado com EXPB e EXPL, o que sugere que estas duas últimas subfamílias possuem uma relação mais estreita entre si, o que de fato é corroborado pelas análises evolutivas.

Figura 22- Análise de motivos da subfamília EXPA em cana-de-açúcar.

Fonte: Arquivo pessoal

Figura 23- Análise de motivos da subfamília EXPB em cana-de-açúcar.

Fonte: Arquivo pessoal

Na subfamília EXPL de cana-de-açúcar também foi possível notar, além da inserção e substituição de alguns motivos, que a sequência correspondente ao motivo 2 foi duplicada em *ShEXPL2*, *ShEXPL3* e *ShEXPL4* e triplicada em *ShEXPL1* (Figura 24). Este evento de repetição não foi observado na mesma análise em sorgo (Apêndice E), o que pode sugerir características próprias de EXPL de cana-de-açúcar. Este resultado pode estar relacionado com o processo de evolução e diferenciação das expansinas-like dentro da espécie, como foi observado por Lu *et al.* (2016) em sua análise de Expansinas de tomate.

Figura 24- Análise de motivos da subfamília EXPL em cana-de-açúcar.

Fonte: Arquivo pessoal

A partir desta análise da estrutura das sequências de Expansinas das espécies estudadas foi possível distinguir a diferença entre o padrão de distribuição de motivos e a estrutura de éxons e introns entre as três subfamílias EXPA, EXPB e EXPL, assim como observar o nível de conservação dos genes de uma mesma subfamília, o que permite inferir que grande parte destes genes podem ser resultados de eventos de duplicação segmental e tandem no decorrer da evolução das plantas e que algumas destas sequências podem apresentar uma certa redundância de função dentro da planta devido à alta similaridade dentro de um mesmo grupo (LU et al., 2016; ZHU et al., 2014a)

Com base nestes dados também se observou que grande parte das Expansinas até o momento indicadas como de possível interesse biotecnológico neste trabalho (*ShEXPA14*, *ShEXPB4* e *ShEXPB22*) possuem em sua estrutura o peptídeo sinal e os domínios conservados HFD, GGACG e TWYG, que podem ser estruturas de grande importância para a função destas proteínas. As exceções foram os transcritos *ShEXPA24*, por esta sequência se mostrar incompleta e *ShEXPL3*, devido à dificuldade de se definir a região do peptídeo sinal e pela ausência dos três motivos citados acima nesta subfamília

5.4 ANÁLISE DE EXPRESSÃO *IN SILICO* DE EXPANSINAS EM CANA-DE-AÇÚCAR

Os transcritos de Expansinas das subfamílias EXPA, EXPB e EXPL identificados nas bibliotecas do SUCEST foram analisados quanto ao nível de expressão (Figuras 25, 26 e 27). Foram escolhidos oito tecidos mais propensos à expressão das Expansinas, observando o perfil de expressão destas proteínas na variedade de tecidos avaliados pelo projeto SUCEST, sendo estes: órgãos florais em diferentes fases do desenvolvimento, sementes em fase de desenvolvimento, folhas imaturas e folhas estioladas de plantas crescidas *in vitro*, primeiro e quarto entrenó de planta maduras, epiderme de plantas maduras, raiz de plantas jovens e raiz e ápice da raiz de plantas maduras.

Figura 25- Análise dos dados de expressão dos transcritos de cana-de-açúcar em EXPA em diferentes bibliotecas do Sucest.

Figura 26- Análise dos dados de expressão dos transcritos de cana-de-açúcar em EXPB em diferentes bibliotecas do Sucest.

Figura 27- Análise dos dados de expressão dos transcritos de cana-de-açúcar em EXPL em diferentes bibliotecas do Sucest.

A partir destas análises pode-se notar que a maioria das Expansinas encontradas apresentaram maior expressão em um tecido específico, sendo poucas aquelas que são expressas em vários órgãos ao mesmo tempo. Esse fato corrobora os resultados do trabalho de Zhang *et al.* (2014c) que avaliou a expressão de Expansinas em milho e verificou que grande

parte destas proteínas são expressas em um único tecido específico, como folha, raiz e caule, enquanto poucas podem ser expressas em mais de um órgão simultaneamente, como é o caso do gene *EXPB4* e *EXPB9* de milho, transcrito em raiz, endosperma e embrião.

O trabalho de Zhu *et al.* (2014) obteve resultados semelhantes ao analisar a expressão de Expansinas em soja. Esses autores observaram que a quantidade de genes expressos de forma constitutiva foi de apenas 23 sequências entre as 75 avaliadas, entretanto, seus resultados demonstraram que estas Expansinas estavam envolvidas em múltiplos processos durante o desenvolvimento da soja. Já as demais sequências estudadas exibiam uma expressão preferencial, sendo abundantemente encontradas em um único tipo de tecido. Este fato também foi observado no estudo de Dal Santo *et al.* (2013), que analisando a expressão de Expansinas em uva perceberam que certos genes eram expressas em um único tecido, como os envolvidos especificamente com o processo de iniciação da formação de pelos radiculares, fenômeno também presente no processo de alongamento da raiz de soja e arroz (YU *et al.*, 2011).

Com base nos dados de expressão apresentados também é possível observar alguns transcritos com regulação positiva no entrenó (*ShEXPA1-zm*, *ShEXPA33*, *ShEXPB28*, *ShEXPB21* e *ShEXPB3*), o que pode indicar um maior efeito destas Expansinas no afrouxamento da parede celular deste órgão, tendo assim, potencial para serem usadas como aditivos para aumentar a taxa de sacarificação da biomassa. Dentre as quatro sequências citadas, os dois transcritos de EXPB tiveram uma maior expressão em comparação com as sequências pertencentes à subfamília EXPA, o que pode estar relacionado à maior especificidade que a parede das gramíneas possuem aos genes EXPB. Os transcritos *ShEXPB3* e *ShEXPB25* também apresentam expressão na epiderme, demonstrando assim uma possível ação em um tecido mais lignificado, o qual demonstra maior recalcitrância ao processo de conversão.

Além de possuírem expressão em órgãos de interesse para este estudo, todas as Expansinas citadas acima possuem os dois domínios característicos desta superfamília, assim como os motivos e resíduos conservados. Desta forma, estes transcritos podem representar sequências nucleotídicas de interesse para este estudo, sendo portanto bons candidatos para as análises de validação da expressão gênica por PCR em tempo real.

5.5 ESTRUTURA PREDITA DE EXPANSINAS DE SORGO E CANA-DE-AÇÚCAR

A estrutura tridimensional de uma proteína fornece importantes informações para o entendimento das propriedades funcionais, papéis biológicos e potenciais aplicações biotecnológicas das mesmas. Atualmente, diversos estudos suportam a ideia de que a conformação 3D das proteínas é orientada por um conjunto de estruturas locais recorrentes nas sequências proteicas. Desta forma, estes segmentos proteicos são relevantes na identificação de dobras nativas, tornando possível a predição por similaridade, de modelos tridimensionais de proteínas ainda não caracterizadas por cristalização e difração de raio X (DOPPELT et al., 2007; BLUNDELL et al., 2006).

Outra importante ferramenta para a predição de estruturas terciárias de proteínas é a informação evolutiva entre sequências homólogas, uma vez que pode haver uma pressão seletiva durante a evolução para que estas sequências mantenham sua estrutura e função similares ou que ainda não tenham divergido significativamente. Assim, homólogos filogeneticamente próximos e com relações de identidade suficiente entre as sequências proteicas primárias permite que modelos tridimensionais possam ser construídos com maior confiabilidade, já que estes genes geralmente possuem estruturas terciárias semelhantes (PEI; GRISHIN, 2004).

Assim, tendo como base a importância da estrutura tridimensional na função e atividade das proteínas realizou-se a predição de modelos 3D de Expansinas de sorgo e cana-de-açúcar por meio de programas de bioinformática. A partir da estrutura cristalina de EXPB1 de milho já elucidada experimentalmente por Yennawar *et al.* (2006) (*Protein Data Bank Archive* (PDB): 2hcz) foi possível construir modelos proteicos confiáveis para as sequências homólogas de sorgo e cana-de-açúcar (*Sb*EXPB13 e *Sh*EXPB13 respectivamente) (Figura 28 A e B).

A partir do mesmo modelo estrutural, e utilizando como base o trabalho de Gaete-Eastman *et al* (2015) que realizou a predição da estrutura de uma Expansina A de *Vasconcellea pubescens* (*Vp*EXPA2) a partir da informação estrutural de *Zm*EXPB1, também realizou se a construção do modelo de *Sh*EXPA1 (Figura 28-C).

Figura 28- A) Predição da estrutura tridimensional de EXPB13 de sorgo. B) Predição da estrutura tridimensional de EXPB13 de cana-de-açúcar. C) Predição da estrutura tridimensional de *ShEXPA1* de cana-de-açúcar. Estruturas em espiral indicam alfa-hélice, fitas representam folhas beta e linhas representam alças de conexão. Regiões em azul marcam resíduos polares das proteínas.

Fonte: Arquivo pessoal

No alinhamento realizado pelos programas integrados do *Swiss model* para a construção da figura, ambos os modelos EXPB apresentaram uma identidade de sequência maior que 80% com EXPB1 de milho, utilizada como molde para a obtenção da estrutura predita. Entretanto, a sequência de EXPA demonstrou uma baixa identidade (73 de 244 aminoácidos, cerca de 30%) com *ZmEXPB1*, porém com uma conformação próxima ao encontrado por Gaete-Eastman *et al* (2015). Isso pode ser devido ao fato de que quando o grau de identidade das sequências primárias da proteína molde e da proteína de interesse for maior que 25% em uma sequência maior que 80 resíduos há uma alta probabilidade de que a estrutura tridimensional delas seja semelhante (ANDRADE; FILHO; ALENCASTRO, 2003; VRIEND; SANDER, 1991).

Desta forma, foi detectado uma alta identidade entre a estrutura terciária da sequência molde (*ZmEXPB1*) e seus homólogos em sorgo e cana-de-açúcar (*SbEXPB13* e *ShEXPB13*, respectivamente) por meio do programa Topmatch (SIPPL; WIEDERSTEIN, 2012), o que possibilitou identificar a cadeia de seis folhas β e duas α -hélices curtas no domínio D1 e as oito folhas β presentes no domínio D2 já previamente caracterizadas pelo trabalho de Yennawar *et al.* (2006) (Figura 28). Através do ensaio experimental realizada por estes autores também foram identificadas três pontes de dissulfeto no domínio D1 e seis resíduos conservados de cisteína envolvidas nestas ligações.

Com base nisto se realizou uma análise das sequências de cana e sorgo pelo banco de dados do Prosite (SIGRIST *et al.*, 2013), sendo assim possível identificar estes mesmos resíduos de cisteína em posições conservadas dentro da proteína, além das três ligações dissulfetos presentes em *ZmEXPB1*. Isto demonstra que as conformações tridimensionais das Expansinas mantêm se altamente similar dentro da uma subfamília, mesmo em diferentes espécies (YENNAWAR *et al.* 2006).

Como demonstrado na figura 29, as Expansinas α e β possuem muitas similaridades em sua estrutura, o que corrobora com os dados de Yennawar *et al.* (2006) que observou cisteínas relacionadas à formação da estrutura 3D conservadas entre as duas subfamílias.

Figura 29- Comparação estrutural de *ShEXPA1* e *ShEXPB13*.

Fonte: Arquivo pessoal

Em seu trabalho Gaete-Eastman *et al.* (2015) discute que, de forma geral, as duas principais diferenças nas sequências de EXPA e EXPB estão relacionadas às α e β inserções, que podem estar envolvidas na perda de uma pequena α -hélice em *VpEXPA2* em comparação com *ZmEXPB1* e a presença de um loop mais longo em *VpEXPA2* do que é encontrado em *ZmEXPB1*. Dados similares também foram observados neste trabalho ao se comparar *ShEXPA1*

e *ShEXPB13*, podendo ser identificado o loop com maior tamanho na estrutura de *ShEXPA1* e perda de uma das α -hélices (Figura 29). Essas pequenas diferenças estruturais presentes em Expansinas de mais de um tipo de espécie vegetal pode indicar uma das razões pela qual EXPAs são mais eficazes em paredes de eudicotiledôneas e EXPB em paredes de gramíneas (GAETE-EASTMAN; MORALES-QUINTANA; HERRERA, 2015; SAMPEDRO et al., 2015; SEKI et al., 2015).

5.6 EXTRAÇÃO DE RNA E VALIDAÇÃO DOS INICIADORES PARA ANÁLISES DE PCR EM TEMPO REAL

A extração do RNA dos entrenós 1, 3, 5, 7 e 9 dos híbridos 140 e 89 de cana-de-açúcar, assim como raiz e folha do híbrido 140 foram feitas em duplicatas biológicas e tiveram sua concentração e as relações 260/230nm e 260/280nm avaliadas por espectrofotometria (Tabelas 8 e 9). O comprimento de onda 260nm é absorvido pelos ácidos nucleicos e assim usado para indicar a concentração de RNA total na amostra. Já as relações 260/230nm e 260/280nm são utilizadas para determinar o “grau de pureza” do RNA, sendo o valor esperado destas razões em torno de 1,8 e 2,1. Contaminantes como proteínas, fenóis e outros componentes absorvem fortemente a faixa de 280nm, tornando esta razão mais baixa, enquanto a presença de outros contaminantes, como por exemplo sais, é indicado por um baixo valor de 260/230nm (THERMOFISHER SCIENTIFIC, 2016).

Levando em conta que técnicas sensíveis de biologia molecular, como construção de bibliotecas de cDNA e PCR em Tempo Real, requerem RNA em quantidades adequadas e em boa qualidade (CARDILLO et al., 2006), buscou-se isolar amostras com estas características. Com base nisto, as amostras extraídas dos dois indivíduos do híbrido 140 (nomeados 140.1 e 140.2) entrenós 3, 5 e 7 apresentaram valores dentro do indicativo de ácidos nucleicos puros (Tabela 8). O mesmo ocorreu com o híbrido 89, indivíduo 89.1 entrenó 3 e 89.2 entrenó 9 (Tabela 9).

Apesar das demais amostras apresentarem relações com valor menor do que o esperado, com exceção da amostra de raiz do híbrido 140.1, estes pequenos desvios, no entanto, podem não interferir em reações posteriores.

Tabela 8- Concentração de RNA total e valor das razões 260/230nm e 260/280nm do córtex em diferentes estágios de amadurecimento, raiz e folha das duplicatas biológicas do híbrido 140.

Amostra	Concentração (ng/ μ L)	A260/230	A260/280
140.1 - Entrenó 1	55,97	1,346	2,189
140.2 - Entrenó 1	169,9	1,394	1,996
140.1 - Entrenó 3	159	1,897	1,915
140.2 - Entrenó 3	169,8	1,925	1,961
140.1 - Entrenó 5	155,5	1,844	1,996
140.2 - Entrenó 5	74,9	2,001	1,966
140.1 - Entrenó 7	246,9	2,149	1,886
140.2 - Entrenó 7	256	1,739	1,916
140.1 - Entrenó 9	105	1,678	1,953
140.2 - Entrenó 9	125,7	1,962	1,825
140.1 - Folha	1670,5	1,639	1,923
140.2 - Folha	1131,9	1,677	1,941
140.1 - Raiz	210,9	1,069	2,018
140.2 - Raiz	205,4	1,478	1,949

Fonte: Arquivo pessoal

Tabela 9- Concentração de RNA total e valor das razões 260/230nm e 260/280nm do córtex em diferentes estágios de amadurecimento das duplicatas biológicas do híbrido 89.

Amostra	Concentração (ng/ μ L)	A260/230	A260/280
89.1 - Entrenó 1	235,1	1,456	1,987
89.2 - Entrenó 1	120,4	1,228	1,993
89.1 - Entrenó 3	121,1	2,089	2,006
89.2 - Entrenó 3	254,7	1,451	2,076
89.1 - Entrenó 5	151,4	1,781	1,946
89.2 - Entrenó 5	232,2	1,725	1,984
89.1 - Entrenó 7	91,1	1,63	1,968
89.2 - Entrenó 7	146,7	2,339	1,922
89.1 - Entrenó 9	119,9	1,339	1,949
89.2 - Entrenó 9	148,3	1,811	1,944

Fonte: Arquivo pessoal

Em seguida, o RNA extraído foi avaliado quanto a sua integridade por eletroforese em gel de agarose (Figura 30 A). A partir disto foi possível observar que a maioria das amostras apresentou as duas bandas características do RNA total em eucariotos, sendo estas compostas pelas partes de RNA ribossomal 28S e 18S (ZHAO et al., 2013), indicando com isso a integridade destes ácidos nucleicos. Entretanto, as amostras 89.2 entrenó 9, 140.1 entrenó 9 e 3 e 140.2 entrenó 5 mostraram se degradadas, o que pode ter acontecido durante o processo de extração ou manipulação das mesma. Já as duas amostras de folha dos indivíduos 140.1 e 140.2, devido à grande quantidade de massa de ácidos nucleicos obtida durante a extração, não foi possível visualizar as duas bandas características da integridade do RNA. Por esse motivo foi realizado

uma diluição e então uma nova avaliação por eletroforese destas amostras (Figura 30 B), o que permitiu observar que a amostra 140.2 de folha possui as bandas 28S e 18S, demonstrando sua integridade, porém a amostra 140.1 de folha se mostrou realmente degradada.

A partir destes resultados foi possível identificar as amostras adequadas para serem utilizadas na técnica de PCR em tempo real. Já as amostras que se comprovaram degradadas pela eletroforese precisarão ser maceradas e ter seu RNA total extraído e avaliado novamente.

Figura 30- (A) Eletroforese em gel de agarose das amostras de RNA extraídas dos indivíduos 140.1, 140.2, 89.1 e 89.2 entrenós 1, 3, 5 e 9 e amostras de folha e raiz dos indivíduos 140.1 e 140.2. (B) Diluição das amostras de folha dos indivíduos 140.1 e 140.2

Fonte: Arquivo pessoal

Em seguida, a especificidade dos iniciadores desenhados para genes da subfamília EXPB em cana-de-açúcar foram validados por PCR tradicional. Para estas análises iniciais da subfamília EXPB se utilizou genes com possível interesse biotecnológico de Expansinas de cana-de-açúcar, sendo estes *EXPB6*, *EXPB27*, *EXPB28*, *EXPB4*, *EXPB12*, *EXPB21*, *EXPB25*, *EXPB3* e *EXPB23*. Os produtos amplificados foram avaliados por eletroforese de ácidos nucleicos (Figura 31).

Figura 31- Avaliação da especificidade dos primers da subfamília de EXPB de cana-de-açúcar

Fonte: Arquivo pessoal. *ShEXPB6*= 1; *ShEXPB27*=2; *ShEXPB21*=3; *ShEXPB12*=4; *ShEXPB4*=5.1 (65°C) e 5.2 (62°C); *ShEXPB28*=6.1 (65°C) e 6.2 (62°C) e *ShEXPB25*=7. C1, C2, C3, C4, C5, C6 e C7 = respectivos controles negativos.

Na figura 31 é possível observar que os iniciadores de *ShEXPB6*, *ShEXPB27*, *ShEXPB21*, *ShEXPB12*, *ShEXPB4*, *ShEXPB28* e *ShEXPB25* representados pelos números 1, 2, 3, 4, 5.1, 5.2, 6.1, 6.2 e 7 respectivamente, mostraram-se específicos com um único produto amplificado. As bandas encontradas estão dentro do tamanho de amplificação esperado, entre 100 e 150 bases, sendo que, no tamanho aproximado de 50 bases, pode se perceber as bandas formadas pelos *primers* não utilizados na reação. Adicionalmente, as reações para os genes *ShEXPB3* e *ShEXPB23* não apresentaram produtos amplificadas (dados não mostrados).

Uma vez que testes indicaram produtos inespecíficos nas reações com os iniciadores de *ShEXPB4* e *ShEXPB28*, as reações foram ajustadas com o aumento da temperatura de anelamento de 60°C, inicialmente utilizada para todos os genes. Na figura 31, a *lane* 5.1 e 5.2 corresponde aos produtos de *ShEXPB4* à temperatura de 65°C e 62°C, respectivamente. Enquanto a *lane* 6.1 e 6.2 corresponde ao produto de amplificação do gene *ShEXPB28* à temperatura de 65°C e 62°C (Figura 31). O aumento da temperatura na etapa de anelamento melhorou a especificidade das reações, resultando em apenas uma banda de amplificação. No entanto, as bandas formadas à 62°C mostraram se mais fortes, o que demonstra que esta temperatura é a mais indicada para as análises futuras.

As amostras de RNA extraídas com sucesso e as condições ajustadas de cada iniciador de genes da subfamília EXPB serão utilizadas nas análises de expressão gênica por PCR em Tempo Real.

6- CONCLUSÃO

Os resultados obtidos no presente trabalho permitiram a prospecção de genes de Expansinas nas espécies de gramíneas próximas à cana-de-açúcar, assim como em diferentes bancos de transcritos e do genoma parcial de cana analisado. Por meio da construção de árvores filogenéticas se analisou as relações de homologia entre as sequências encontradas e com base nisso se identificou aquelas de maior interesse em cana-de-açúcar. Dessa forma, as sequências nucleotídicas *ShEXPA14*, *ShEXPA24*, *ShEXPB22* e *ShEXPL3* de cana foram considerados de relevância para esse estudo, uma vez que são ortólogos às sequências proteicas de arroz que possuem bons resultados na promoção da sacarificação, segundo o artigo de Seki *et al.* (2014). Adicionalmente, a análise da expressão de transcritos presentes em bibliotecas do SUCEST também permitiu selecionar as sequências nucleotídicas que podem atuar no aumento da taxa de sacarificação enzimática, sendo estas *ShEXPA1*, *ShEXPA33*, *ShEXPB28*, *ShEXPB3*, *ShEXPB21* e *ShEXPB25*. Já *ShEXPA24* e *ShEXPA14* também apresentam interesse por serem filogeneticamente relacionadas à proteína de arroz envolvida no aumento da biomassa vegetal (BUCHANAN *et al.*, 2005; MA *et al.*, 2013). *ShEXPB4* possui potencial aplicação biotecnológica em processos diversos, como na resposta à desidratação e alta salinidade.

As análises dos domínios e motivos conservados das Expansinas de sorgo e cana-de-açúcar, assim como a construção da estrutura predita de *ShEXPB13*, *SbEXPB13* e *ShEXPB1*, permitiram um estudo mais aprofundado das subfamílias EXPA, EXPB e EXPL. A partir dos resultados obtidos foi possível concluir que alguns motivos conservados estão presentes em diferentes espécies de plantas e podem estar envolvidas no mecanismo de ação das Expansinas vegetais, além disto, algumas pequenas diferenças na estrutura e motivos das sequências de EXPA e EXPB podem contribuir para a maior especificidade desta última à parede celular das monocotiledôneas.

Em adição, a extração de RNA dos tecidos coletados de entrenós e medulas em diferentes estágios do desenvolvimento da planta, assim como raiz e folhas de plântulas de cana-de-açúcar possibilitou a construção de bancos de cDNA que podem ser usados em futuras análises de PCR em Tempo Real para avaliar a taxa de expressão das Expansinas de interesse.

7- PERSPECTIVAS

Posteriormente, um estudo aprofundado das características e atividade das Expansinas em cana-de-açúcar, assim como seu sinergismo com celulasas para aumentar a eficiência da hidrólise enzimática do bagaço para posterior produção de etanol contaria com as seguintes propostas:

- Análise de expressão por PCR em Tempo Real;
- Clonagem e expressão heteróloga de Expansinas de cana-de-açúcar com potencial aplicação biotecnológica;
- Determinação da estrutura tridimensional de Expansinas de interesse por cristalinização;
- Avaliação do sinergismo em ensaios de hidrólise enzimática e determinação da taxa de conversão de biomassa em açúcares fermentescíveis.

REFERÊNCIAS

- ARAÚJO, N.D. et al., A era da bioinformática : seu potencial e suas implicações para as ciências da saúde. **Faculdade Integrada de Patos (FIP)**, v. 30, p. 143–148, 2008.
- ANDRADE, O.; FILHO, S.; ALENCASTRO, B. DE., Modelagem de proteínas por homologia, **Química Nova**, v. 26, n. 2, p. 253–259, 2003.
- ASAKAWA, A. et al. Comparison of choline acetate ionic liquid pretreatment with various pretreatments for enhancing the enzymatic saccharification of sugarcane bagasse. **Industrial Crops and Products**, v. 71, p. 147–152, 2015.
- BAILEY, T. L. et al. MEME: Discovering and analyzing DNA and protein sequence motifs. **Nucleic Acids Research**, v. 34, 2006.
- BAILEY, T. L. et al. MEME Suite: tools for motif discovery and searching. **Nucleic Acids Research**, v. 37, n. SUPPL. 2, 2009.
- BALESTRINI, R.; COSGROVE, D. J.; BONFANTE, P. Differential location of alpha-expansin proteins during the accommodation of root cells to an arbuscular mycorrhizal fungus. **Planta**, v. 220, n. 6, p. 889–99, 2005.
- BLUNDELL, T. L. et al. Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. **Philos Trans R Soc Lond B Biol Sci**, v. 361, n. 1467, p. 413–423, 2006.
- BUCHANAN, C. D. et al. Sorghum bicolor's transcriptome response to dehydration, high salinity and ABA. **Plant Molecular Biology**, v. 58, n. 5, p. 699–720, 2005.
- CARDOSO-SILVA, C. B. et al. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. **PloS one**, v. 9, n. 2, p. e88462, 2014.
- CAREY, R. E.; HEPLER, N. K.; COSGROVE, D. J. Selaginella moellendorffii has a reduced and highly conserved expansin superfamily with genes more closely related to angiosperms than to bryophytes. **BMC plant biology**, v. 13, n. 1, p. 4, 2013.
- CARPITA, N. C. Structure and biogenesis of the cell walls of grasses. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 47, n. 1, p. 445–476, 1996.
- CARROLL, A.; SOMERVILLE, C. Cellulosic biofuels. **Annual review of plant biology**, v. 60, p. 165–82, 2009.
- CHIARA, M.; HORNER, D. S.; SPADA, A. De novo assembly of the transcriptome of the non-model plant *Streptocarpus rexii* employing a novel heuristic to recover locus-specific transcript clusters. **PloS one**, v. 8, n. 12, p. e80961, 2013.
- CHO, H. T.; COSGROVE, D. J. Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 17, p. 9783–9788, 2000.
- CHO, H. T.; KENDE, H. Expression of expansin genes is correlated with growth in deepwater rice. **The Plant cell**, v. 9, n. 9, p. 1661–1671, 1997.
- COSGROVE, D. J. Loosening of plant cell walls by expansins. **Nature**, v. 407, n. 6802, p. 321–326, 2000a.
- COSGROVE, D. J. New genes and new biological roles for expansins. **Growth and Development**, p. 73–

78, 2000b.

COSGROVE, D. J. Plant expansins: diversity and interactions with plant cell walls. **Current Opinion in Plant Biology**, v. 25, p. 162–172, 2015.

COSGROVE, D. J.; BEDINGER, P.; DURACHKO, D. M. Group I allergens of grass pollen as cell wall-loosening agents. **Proceedings of the National Academy of Sciences**, v. 94, n. 12, p. 6559–6564, 1997.

DAL SANTO, S. et al. Genome-wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics. **PloS one**, v. 8, n. 4, p. e62206, 2013.

DOPPELT, O. et al. Functionnal annotation strategy for protein structures. **Bioinformatics**, v. 1, n. 9, p. 357–359, 2007.

DRABIK, D.; CIAIAN, P.; POKRIVČÁK, J. The effect of ethanol policies on the vertical price transmission in corn and food markets. **Energy Economics**, v. 55, p. 189–199, 2016.

EDDY, S. R. A new generation of homology search tools based on probabilistic inference. **Genome informatics. International Conference on Genome Informatics**, v. 23, n. 1, p. 205–211, 2009.

GAETE-EASTMAN, C.; MORALES-QUINTANA, L.; HERRERA, R. In-silico analysis of the structure and binding site features of an α -expansin protein from mountain papaya fruit (VpEXPA2), through molecular modeling , docking , and dynamics simulation studies. p. 1–12, 2015.

GEORGELIS, N.; YENNAWAR, N. H.; COSGROVE, D. J. Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 37, p. 14830–14835, 2012.

GIORDANO, W.; HIRSCH, A. M. The expression of MaEXP1, a Melilotus alba expansin gene, is upregulated during the sweetclover-Sinorhizobium meliloti interaction. **Molecular plant-microbe interactions : MPMI**, v. 17, n. 6, p. 613–22, 2004.

GOLDEMBERG, J. The brazilian biofuels industry. **Biotechnology for Biofuels**, v. 1, n. 1, p. 6, 2008.

GOUY, M.; GUINDON, S.; GASCUEL, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Molecular biology and evolution**, v. 27, n. 2, p. 221–224, 2010.

GRATIVOL, C. et al. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. **Plant Journal**, v. 79, n. 1, p. 162–172, 2014.

GRAY-MITSUMUNE, M. et al. Expansins abundant in secondary xylem belong to subgroup A of the alpha-expansin gene family. **Plant physiology**, v. 135, n. 3, p. 1552–1564, 2004.

GUINDON, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. **Systematic Biology**, v. 59, n. 3, p. 307–321, 2010.

HALL, T. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NTNucleic Acids Symposium Series**, 1999. Disponível em: <http://jwbrown.mbio.ncsu.edu/JWB/papers/1999Hall1.pdf> - acesso em: 10 ago/2016

HAN, Y. et al. Over-expression of TaEXPB23 , a wheat expansin gene , improves oxidative stress tolerance in transgenic tobacco plants. **Journal of Plant Physiology**, v. 173, p. 62–71, 2015.

HAN, Y. Y. et al. Characterization of a wheat (Triticum aestivum L.) expansin gene, TaEXPB23, involved in the abiotic stress response and phytohormone regulation. **Plant Physiology and Biochemistry**, v. 54, p. 49–58, 2012.

HAN, Y. Y. et al. The involvement of expansins in responses to phosphorus availability in wheat, and its

potentials in improving phosphorus efficiency of plants. **Plant Physiology and Biochemistry**, v. 78, p. 53–62, 2014.

IEA, Technology Roadmap: Biofuels for Transport; **International Energy Agency**, 2011

JAKOB, K.; ZHOU, F.; PATERSON, A. H. Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. **In Vitro Cellular and Developmental Biology - Plant**, v. 45, n. 3, p. 291–305, 2009.

JONES, L.; MCQUEEN-MASON, S. A role for expansins in dehydration and rehydration of the resurrection plant *Craterostigma plantagineum*. **FEBS Letters**, v. 559, n. 1-3, p. 61–65, 2004.

KERFF, F. et al. Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 44, p. 16876–16881, 2008.

KOONIN, E. V. Orthologs, paralogs, and evolutionary genomics. **Annual Review of Genetics**, v. 39, n. 1, p. 309–338, 2005.

KRISHNAMURTHY, P. et al. Genome-wide analysis of the expansin gene superfamily reveals *Brassica rapa*-specific evolutionary dynamics upon whole genome triplication. **Molecular Genetics and Genomics**, v. 290, n. 2, p. 521–530, 2014.

LARKIN, M. A. et al. Clustal W and Clustal X version 2.0. **Bioinformatics**, v. 23, n. 21, p. 2947–2948, 2007.

LAURITO-FRIEND, D. F. et al. Sugarcane hybrids with original low lignin contents and high field productivity are useful to reach high glucose yields from bagasse. **Biomass and Bioenergy**, v. 75, p. 65–74, 2015.

LEE, D.-S. et al. Characterization of a new α -L: arabinofuranosidase from *Penicillium* sp. LYG 0704, and their application in lignocelluloses degradation. **Molecular Biotechnology**, v. 49, n. 3, p. 229–239, 2011.

LEE, D.-S. et al. Rapid saccharification for production of cellulosic biofuels. **Bioresource Technology**, v. 158, p. 239–247, 2014.

LEE, Y.; CHOI, D.; KENDE, H. Expansins: ever-expanding numbers and functions. **Current Opinion in Plant Biology**, v. 4, n. 6, p. 527–532, 2001.

LI, L.-C. et al. Purification and characterization of four beta-expansins (*Zea m 1* isoforms) from maize pollen. **Plant physiology**, v. 132, n. 4, p. 2073–2085, 2003.

LI, Y. et al. Plant expansins are a complex multigene family with an ancient evolutionary origin. **Plant physiology**, v. 128, n. 3, p. 854–864, 2002.

LISBOA, C. C. et al. Bioethanol production from sugarcane and emissions of greenhouse gases - known and unknowns. **GCB Bioenergy**, v. 3, n. 4, p. 277–292, 2011.

LIU, X. et al. Heterologous expression and functional characterization of a novel cellulose-disruptive protein LeEXP2 from *Lycopersicon esculentum*. **Journal of Biotechnology**, v. 186, p. 148–155, 2014a.

LIU, X. et al. Heterologous expression and functional characterization of a novel cellulose-disruptive protein Le EXP2 from *Lycopersicon esculentum*. **Journal of Biotechnology**, v. 186, p. 148–155, 2014b.

LIU, X.; MA, Y.; ZHANG, M. Research advances in expansins and expansion-like proteins involved in lignocellulose degradation. **Biotechnology Letters**, 2015.

LIU, Y. et al. the α - and β -expansin and xyloglucan endotransglucosylase / hydrolase gene families of wheat : molecular cloning , gene expression , and EST data mining, **BMC Genomics**. v. 90, p. 516–529,

2007.

LU, Y. et al. Genome - wide identification and expression analysis of the expansin gene family in tomato. **Molecular Genetics and Genomics**, v. 291, n. 2, p. 597–608, 2016.

MA, N. et al. Overexpression of OsEXPA8, a root-specific gene, improves rice growth and root system architecture by facilitating cell extension. **PLoS ONE**, v. 8, n. 10, 2013.

MASARIN, F. et al. Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. **Biotechnology for Biofuels**, v. 4, n. 1, p. 55, 2011.

MASUM, B. M. et al. Effect of ethanol–gasoline blend on NOx emission in SI engine. **Renewable and Sustainable Energy Reviews**, v. 24, p. 209–222, 2013.

MCCORMICK, N.; HOWARD, G. Beating back biofuel crop invasions: guidelines on managing the invasive risk of biofuel developments. **Renewable Energy**, v. 49, p. 263–266, 2013.

MCQUEEN-MASON, S.; COSGROVE, D. J. Disruption of hydrogen bonding between plant cell wall polymers by proteins that induce wall extension. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 14, p. 6574–6578, 1994.

MCQUEEN-MASON, S.; DURACHKO, D. M.; COSGROVE, D. J. Two endogenous proteins that induce cell wall extension in plants. **The Plant cell**, v. 4, n. November, p. 1425–1433, 1992.

NIKOLAIDIS, N.; DORAN, N.; COSGROVE, D. J. Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. **Molecular Biology and Evolution**, v. 31, n. 2, p. 376–386, 2014.

NISHIYAMA, M. Y. et al. Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid and ancestor genotypes. **PloS one**, v. 9, n. 9, p. e107351, 2014.

OJEDA, K. et al. Evaluation of technological alternatives for process integration of sugarcane bagasse for sustainable biofuels production—Part 1. **Chemical Engineering Research and Design**, v. 89, n. 3, p. 270–279, 2011.

OLIVEIRA, M. E. D.; MORAES, S. O.; Modeling approaches for agricultural N₂O fluxes from large scale areas: a case for sugarcane crops in the state of São Paulo – Brazil; **Agricultural Systems**, v. 150, January 2017

PEI, J.; GRISHIN, N. V. Combining evolutionary and structural information for local protein structure prediction. **Proteins: Structure, Function and Genetics**, v. 56, n. 4, p. 782–794, 2004

RAVEN, P. H.; EVERT, R. F.; EICHHORN, S. E. **Biologia vegetal**. 5^a edição, Rio de Janeiro, Guanabara, 1996.

RIZZO, J.; ROUCHKA, E. C. Review of phylogenetic tree construction, **University of Louisville bioinformatics laboratory technical report series**. Louisville, 2007.

SAMPEDRO, J. et al. Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. **Plant Journal**, v. 44, n. 3, p. 409–419, 2005.

SAMPEDRO, J. et al. Evolutionary divergence of β -expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits. **The Plant Journal**, v. 81, n. 1, p. 108–120, 2015.

SAMPEDRO, J.; CAREY, R. E.; COSGROVE, D. J. Genome histories clarify evolution of the expansin superfamily: New insights from the poplar genome and pine ESTs. **Journal of Plant Research**, v. 119, n. 1, p. 11–21, 2006.

SAMPEDRO, J.; COSGROVE, D. J. The expansin superfamily. **Genome biology**, v. 6, n. 12, p. 242, 2005a.

- SAMPEDRO, J.; COSGROVE, D. J. Protein family review The expansin superfamily, **Genome biology**, n. 3, p. 1–11, 2005b.
- SANTOS, F. A. et al. POTENCIAL DA PALHA DE CANA-DE-AÇÚCAR PARA PRODUÇÃO DE ETANOL. **Química Nova**, v. 35, n. 5, p. 1004–1010, 2012.
- SEKI, Y. et al. Promotion of crystalline cellulose degradation by expansins from *Oryza sativa*. **Planta**, v. 241, n. 1, p. 83–93, 2014.
- SEKI, Y. et al. Promotion of crystalline cellulose degradation by expansins from *Oryza sativa*. p. 83–93, 2015.
- SIGRIST, C. J. A. et al. New and continuing developments at PROSITE. **Nucleic Acids Research**, v. 41, n. D1, 2013.
- SIMS, R. E. H. et al. Bioresource technology an overview of second generation biofuel technologies. **Bioresource Technology**, v. 101, n. 6, p. 1570–1580, 2010.
- SINDHU, R. et al. Bioconversion of sugarcane crop residue for value added products - an overview. **Renewable Energy**, 2016.
- SIPPL, M. J.; WIEDERSTEIN, M. Detection of spatial correlations in protein structures and molecular complexes. **Structure**, v. 20, n. 4, p. 718–728, 2012.
- SKOLRUD, T. D. et al. The role of federal renewable fuel standards and market structure on the growth of the cellulosic biofuel sector. **Energy Economics**, v. 58, p. 141–151, 2016.
- SMITH, A. L. et al. Second generation biofuels and bioinvasions: an evaluation of invasive risks and policy responses in the United States and Canada. **Renewable and Sustainable Energy Reviews**, v. 27, p. 30–42, 2013.
- SUN, Y.; CHENG, J. Hydrolysis of lignocellulosic materials for ethanol production: a review. **Bioresource Technology**, v. 83, n. 1, p. 1–11, 2002.
- TAMURA, K. et al. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, v. 30, n. 12, p. 2725–2729, 2013.
- THERMOFISHER SCIENTIFIC. **NanoDrop 2000/2000c Spectrophotometers V1.0 Manual**. Disponível em: <<https://www.thermofisher.com/order/catalog/product/ND-2000?CID=search-product>>. Acesso em: 18/nov. 2016.
- VAIN, P. Brachypodium as a model system for grass research. **Journal of Cereal Science**, v. 54, n. 1, p. 1–7, 2011.
- VALENCIA, M. J.; CARDONA, C. A. The Colombian biofuel supply chains: the assessment of current and promising scenarios based on environmental goals. **Energy Policy**, v. 67, p. 232–242, 2014.
- VANNERUM, K. et al. Transcriptional analysis of cell growth and morphogenesis in the unicellular green alga *Micrasterias* (Streptophyta), with emphasis on the role of expansin. **BMC Plant Biology**, v. 11, n. 1, p. 128, 2011.
- VETTORE, A. L. et al. The libraries that made SUCEST. **Genetics and Molecular Biology**, v. 24, n. 1-4, p. 1–7, 2001.
- VETTORE, A. L. et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, v. 13, n. 12, p. 2725–2735, 2003.
- VICENTINI, R. et al. Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. **PLoS ONE**, v. 10, n. 8, p. 1–19, 2015.

- VRIEND, G.; SANDER, C. Detection of common three-dimensional substructures in proteins. **Proteins**, v. 11, n. 1, p. 52–8, 1991.
- WANG, K. et al. Characterization of seed proteome in *Brachypodium distachyon*. **Journal of Cereal Science**, v. 52, n. 2, p. 177–186, 2010.
- WU, Y.; MEELEY, R. B.; COSGROVE, D. J. Analysis and expression of the α -expansin and β -expansin gene families in maize. **Plant Physiology**, v. 126, n. 1, p. 222–232, 2001.
- XU, B. et al. A cotton BURP domain protein interacts with α -expansin and their co-expression promotes plant growth and fruit production. **Molecular Plant**, v. 6, n. 3, p. 945–958, 2013.
- XU, J. et al. Identification and characterization of an expansin gene AsEXP1 associated with heat tolerance in C3 *Agrostis* grass species. **Journal of Experimental Botany**, v. 58, n. 13, p. 3789–3796, 2007.
- XU, Q. et al. Transgenic tobacco plants overexpressing a Grass PpEXP1 gene exhibit enhanced tolerance to heat stress. **PLoS ONE**, v. 9, n. 7, p. e100792, 2014.
- YENNAWAR, N. H. et al. Crystal structure and activities of EXPB1 (*Zea m 1*), a beta-expansin and group-1 pollen allergen from maize. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 40, p. 14664–14671, 2006.
- YU, Z. et al. Root hair-specific expansins modulate root hair elongation in rice. **Plant Journal**, v. 66, n. 5, p. 725–734, 2011.
- ZENONI, S. et al. Downregulation of the petunia hybrida α -expansin gene PhEXP1 reduces the amount of crystalline cellulose in cell walls and leads to phenotypic changes in petal limbs. **Plant Cell**, v. 16, n. 2, p. 295–308, 2004.
- ZENONI, S. et al. Overexpression of PhEXPA1 increases cell size, modifies cell wall polymer composition and affects the timing of axillary meristem development in *Petunia hybrida*. **New Phytologist**, v. 191, n. 3, p. 662–677, 2011.
- ZHANG, K.; PEI, Z.; WANG, D. Organic solvent pretreatment of lignocellulosic biomass for biofuels and biochemicals: a review. **Bioresource Technology**, 2015.
- ZHANG, Q. et al. Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. **BMC plant biology**, v. 14, p. 27, 2014a.
- ZHANG, S. et al. A genome-wide analysis of the expansin genes in *Malus × Domestica*. **Molecular Genetics and Genomics**, v. 289, n. 2, p. 225–236, 2014b.
- ZHANG, W. et al. Genome-wide identification and characterization of maize expansin genes expressed in endosperm. **Molecular Genetics and Genomics**, v. 289, n. 6, p. 1061–1074, 2014c.
- ZHAO, Y. E. et al. Secondary structure prediction for complete rDNA sequences (18S, 5.8S, and 28S rDNA) of *Demodex folliculorum*, and comparison of divergent domains structures across *Acari*. **Experimental Parasitology**, v. 135, n. 2, p. 370–381, 2013.
- ZHU, Y. et al. Soybean (*Glycine max*) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. **BMC plant biology**, v. 14, n. 1, p. 93, 2014a.
- ZHU, Y. et al. Soybean (*Glycine max*) expansin gene superfamily origins : segmental and tandem duplication events followed by divergent selection among subfamilies. v. 14, n. 1, p. 1–19, 2014b.

APÊNDICES

APÊNDICE A

Número de acesso das sequências de α -expansinas em arroz, milho, sorgo, *Brachypodium distachyon* e *Arabidopsis thaliana*, seguido da nomeação da sequência utilizada para a construção da árvore filogenética.

Arroz		Milho		Sorgo		<i>Brachypodium distachyon</i>		<i>Arabidopsis thaliana</i>	
Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação
Os04g15840	OsEXPA1	GRMZM2G339122	ZmEXPA1	001G033300.1	SbEXPA1	Bradi1g03640.1	BdEXPA1	At1g69530	AtEXPA1
Os01g60770	OsEXPA2	GRMZM2G105844	ZmEXPA2	001G237800.1	SbEXPA2	Bradi1g35830.1	BdEXPA2	At5g05290	AtEXPA2
Os05g19570	OsEXPA3	GRMZM2G074585	ZmEXPA3	001G237900.1	SbEXPA3	Bradi1g51990.1	BdEXPA3	At2g37640	AtEXPA3
Os05g39990	OsEXPA4	GRMZM2G106899	ZmEXPA3-a	001G238000.1	SbEXPA4	Bradi1g61190.1	BdEXPA4	At2g39700	AtEXPA4
Os02g51040	OsEXPA5	GRMZM2G368886	ZmEXPA4	001G238100.1	SbEXPA5	Bradi1g74710.1	BdEXPA5	AT3G29030	AtEXPA5
Os03g21820	OsEXPA6	GRMZM2G361064	ZmEXPA5	001G238200.1	SbEXPA6	Bradi1g74720.1	BdEXPA6	AT2G28950	AtEXPA6
Os03g60720	OsEXPA7	GRMZM2G445169	ZmEXPA6	001G238300.1	SbEXPA7	Bradi1g74740.1	BdEXPA7	AT1G12560	AtEXPA7
Os01g14650	OsEXPA8	GRMZM2G066862	ZmEXPA7	001G238400.1	SbEXPA8	Bradi1g74750.1	BdEXPA8	AT2G40610	AtEXPA8
Os01g14660	OsEXPA9	GRMZM2G127029	ZmEXPA8	001G314600.1	SbEXPA9	Bradi2g08760.1	BdEXPA9	AT5G02260	AtEXPA9
Os04g49410	OsEXPA10	GRMZM2G0871013	ZmEXPA9	001G356400.1	SbEXPA10	Bradi2g08780.1	BdEXPA10	At1g26770	AtEXPA10
Os01g16770	OsEXPA11	GRMZM2G126196	ZmEXPA10	001G499800.1	SbEXPA11	Bradi2g10320.1	BdEXPA11	At1g20190	AtEXPA11
Os03g06000	OsEXPA12	GRMZM2G149605	ZmEXPA11	001G499900.1	SbEXPA12	Bradi2g22290.1	BdEXPA12	At3g15370	AtEXPA12
Os02g16730	OsEXPA13	AC219190.3	ZmEXPA12	002G305800.1	SbEXPA13	Bradi2g31760.1	BdEXPA13	At3g03220	AtEXPA13
Os02g16780	OsEXPA14	GRMZM2G120724	ZmEXPA13	003G059900.1	SbEXPA14	Bradi2g53580.1	BdEXPA14	At5g56320	AtEXPA14
Os03g06020	OsEXPA15	GRMZM2G030531	ZmEXPA14	003G112100.1	SbEXPA15	Bradi3g09930.1	BdEXPA15	At2g03090	AtEXPA15
Os06g41700	OsEXPA16	GRMZM2G105844	ZmEXPA15	003G128800.1	SbEXPA16	Bradi3g09940.1	BdEXPA16	At3g55500	AtEXPA16
Os06g01920	OsEXPA17	GRMZM2G139695	ZmEXPA16	003G444400.1	SbEXPA17	Bradi3g09950.1	BdEXPA17	At4g01630	AtEXPA17
Os03g06040	OsEXPA18	GRMZM2G072121	ZmEXPA17	004G119600.1	SbEXPA18	Bradi3g09960.1	BdEXPA18	At1g62980	AtEXPA18
Os03g06050	OsEXPA19	GRMZM2G173826	ZmEXPA18	004G119800.1	SbEXPA19	Bradi3g09967.1	BdEXPA19	AY843212	AtEXPA19
Os03g06060	OsEXPA20	GRMZM2G035189	ZmEXPA19	004G119900.1	SbEXPA20	Bradi3g09990.1	BdEXPA20	AT4G38210	AtEXPA20
Os03g25990	OsEXPA21	GRMZM2G332412	ZmEXPA20	004G120000.1	SbEXPA21	Bradi3g10100.1	BdEXPA21	AT5G39260	AtEXPA21
Os02g16850	OsEXPA22	GRMZM2G115909	ZmEXPA21	004G121600.1	SbEXPA22	Bradi3g13720.1	BdEXPA22	AT5G39270	AtEXPA22
Os02g16809	OsEXPA23.1	GRMZM2G414779	ZmEXPA22	004G121700.1	SbEXPA23	Bradi3g19500.1	BdEXPA23	AT5G39280	AtEXPA23
Os02g16839	OsEXPA23.2	AC234190.1	ZmEXPA23	004G121800.1	SbEXPA24	Bradi3g27440.1	BdEXPA24	AT5G39310	AtEXPA24
Os02g16800	OsEXPA24	GRMZM2G094523	ZmEXPA24	004G121900.1	SbEXPA25	Bradi3g27450.1	BdEXPA25	AT5G39300	AtEXPA25
Os03g06010	OsEXPA25	AC233901.1	ZmEXPA25	006G031900.1	SbEXPA26	Bradi3g27460.1	BdEXPA26	AT5G39290	AtEXPA26
Os12g36040	OsEXPA26	GRMZM2G073373	ZmEXPA26	006G191700.1	SbEXPA27	Bradi3g27470.1	BdEXPA27		
Os10g30330	OsEXPA27	GRMZM2G069332	ZmEXPA27	007G018000.1	SbEXPA28	Bradi3g32070.1	BdEXPA28		
Os10g30340	OsEXPA28	GRMZM2G359799	ZmEXPA28	007G019900.1	SbEXPA29	Bradi3g43080.1	BdEXPA29		
Os06g50400	OsEXPA29	GRMZM2G088791	ZmEXPA29	007G020100.1	SbEXPA30	Bradi3g59460.1	BdEXPA30		
Os10g39110	OsEXPA30	GRMZM2G146540	ZmEXPA30	007G020200.1	SbEXPA31	Bradi5g04120.1	BdEXPA31		
Os03g31480	OsEXPA31	GRMZM2G019398	ZmEXPA31	007G166500.1	SbEXPA32	Bradi5g19340.2	BdEXPA32		
Os08g44790	OsEXPA32	GRMZM2G303937	ZmEXPA32	009G173700.1	SbEXPA33				
Os05g19600	OsEXPA33	GRMZM2G374248	ZmEXPA33	010G006200.1	SbEXPA34				
		GRMZM2G064174	ZmEXPA34	010G194500.1	SbEXPA35				
		GRMZM2G160314	ZmEXPA35	010G268700.1	SbEXPA36				
		GRMZM2G450546	ZmEXPA36						
		GRMZM2G469689	ZmEXP37						
		GRMZM2G469701	ZmEXP38						

Fonte: Arquivo pessoal

Número de acesso das sequências de β -expansinas em arroz, milho, sorgo, *Brachypodium distachyon* e *Arabidopsis thaliana*, seguido da nomeação da sequência utilizada para a construção da árvore filogenética.

Arroz		Milho		Sorgo		<i>Brachypodium distachyon</i>		<i>Arabidopsis thaliana</i>	
Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação
Os03g01610	OsEXPB1-a	GRMZM2G146551	ZmEXPB1	001G155600.4	SbEXPB1	Bradi1g13787.2	BdEXPB1	AT2G20750	AtEXPB1
Os03g01650	OsEXPB1-b	GRMZM2G146551	ZmEXPB1-a	001G155700.5	SbEXPB2	Bradi1g78120.1	BdEXPB2	AT1G65680	AtEXPB2
Os10g40710	OsEXPB2	GRMZM2G094990	ZmEXPB1a-a	001G300400.1	SbEXPB3	Bradi1g78340.2	BdEXPB3	AT4G28250	AtEXPB3
Os10g40720	OsEXPB3	GRMZM2G021621	ZmEXPB1a-b	001G300500.1	SbEXPB4	Bradi1g78350.1	BdEXPB4	AT2G45110	AtEXPB4
Os10g40730	OsEXPB4	GRMZM2G082520	ZmEXPB1a-c	001G300700.1	SbEXPB5	Bradi2g35007.1	BdEXPB5	AT3G60570	AtEXPB5
Os04g46650	OsEXPB5	GRMZM2G177391	ZmEXPB1a-d	001G300800.1	SbEXPB6	Bradi3g32810.1	BdEXPB6	AY619565	AtEXPB6
Os10g40700	OsEXPB6	GRMZM2G100931	ZmEXPB2	001G300900.2	SbEXPB7	Bradi3g33106.1	BdEXPB7		
Os03g01270	OsEXPB7	GRMZM2G100931	ZmEXPB2-a	001G301000.1	SbEXPB8	Bradi3g33112.1	BdEXPB8		
Os03g01260	OsEXPB8	GRMZM2G169967	ZmEXPB3	001G301300.1	SbEXPB9	Bradi3g33120.1	BdEXPB9		
Os10g40090	OsEXPB9	GRMZM2G078279	ZmEXPB3-a	001G301400.1	SbEXPB10	Bradi3g33130.1	BdEXPB10		
Os03g01640	OsEXPB10	GRMZM2G154178	ZmEXPB4	001G301500.1	SbEXPB11	Bradi3g33140.1	BdEXPB11		
Os02g44108	OsEXPB11	GRMZM2G103672	ZmEXPB4-a	001G301600.1	SbEXPB12	Bradi3g33150.1	BdEXPB12		
Os03g44290	OsEXPB12	GRMZM5G859316	ZmEXPB4-b	001G306200.1	SbEXPB13	Bradi3g33160.1	BdEXPB13		
Os03g01630	OsEXPB13	GRMZM2G013671	ZmEXPB5	001G306200.2	SbEXPB13-a	Bradi3g49850.1	BdEXPB14		
Os02g44106	OsEXPB14	GRMZM2G176595	ZmEXPB6	001G306400.1	SbEXPB14	Bradi3g50730.1	BdEXPB15		
Os04g46630	OsEXPB15	GRMZM2G342246	ZmEXPB7	001G306500.1	SbEXPB15	Bradi3g50740.1	BdEXPB16		
Os02g42650	OsEXPB16	GRMZM2G013002	ZmEXPB8	001G306500.2	SbEXPB15-a	Bradi3g50750.1	BdEXPB17		
Os04g44780	OsEXPB17	GRMZM2G072886	ZmEXPB9	001G539600.1	SbEXPB16	Bradi4g00360.1	BdEXPB18		
Os05g15690	OsEXPB18	GRMZM2G007685	ZmEXPB10	001G539700.1	SbEXPB17	Bradi4g41110.1	BdEXPB19		
		GRMZM2G164785	ZmEXPB10-a	001G539880.1	SbEXPB18	Bradi4g41117.1	BdEXPB20		
		GRMZM2G127106	ZmEXPB10-b	001G539820.1	SbEXPB19	Bradi5g16497.1	BdEXPB21		
		GRMZM2G089699	ZmEXPB11	001G539940.1	SbEXPB19-a	Bradi5g17760.1	BdEXPB22		
		GRMZM2G026956	ZmEXPB12	001G540000.2	SbEXPB20	Bradi5g17770.1	BdEXPB23		
		GRMZM2G148485	ZmEXPB13	001G542100.1	SbEXPB21	Bradi5g17780.1	BdEXPB24		
		GRMZM2G118873	ZmEXPB14	001G542200.1	SbEXPB22				
		GRMZM2G121308	ZmEXPB15	002G124400.1	SbEXPB23				
		GRMZM2G056236	ZmEXPB16	002G124500.1	SbEXPB24				
		GRMZM2G327266	ZmEXPB17	004G191600.1	SbEXPB25				
		GRMZM2G095204	ZmEXPB18	004G227900.1	SbEXPB26				
		GRMZM2G401975	ZmEXPB19	004G294300.1	SbEXPB27				
		GRMZM2G401983	ZmEXPB20	004G294400.1	SbEXPB28				
		GRMZM2G405622	ZmEXPB21	004G294500.1	SbEXPB29				
		GRMZM2G059785	ZmEXPB22	006G171100.1	SbEXPB30				
		GRMZM2G073260	ZmEXPB23	006G171200.1	SbEXPB31				
		GRMZM2G097229	ZmEXPB24	006G171300.1	SbEXPB32				
		GRMZM2G021427	ZmEXPB25	010G120100.1	SbEXPB33				
		GRMZM2G104013	ZmEXPB26	010G121800.1	SbEXPB34				
		GRMZM2G474194	ZmEXPB27						
		GRMZM2G181202	ZmEXPB28						
		GRMZM2G020852	ZmEXPB29						
		GRMZM2G181202	ZmEXPB30						
		GRMZM2G409286	ZmEXPB31						
		GRMZM2G109973	ZmEXPB32						
		GRMZM2G110013	ZmEXPB33						
		AC233881.1	ZmEXPB34						
		GRMZM2G144856	ZmEXPB35						
		GRMZM2G110025	ZmEXPB36						
		GRMZM2G144898	ZmEXPB37						

Fonte: Arquivo pessoal

Número de acesso das sequências de Expansinas-like em arroz, milho, sorgo, *Brachypodium distachyon* e *Arabidopsis thaliana*, seguido da nomeação da sequência utilizada para a construção da árvore filogenética.

Arroz		Milho		Sorgo		<i>Brachypodium distachyon</i>		<i>Arabidopsis thaliana</i>	
Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação	Número de acesso	Nomeação
Os03g04020	OsEXPL1	GRMZM2G114322	ZmEXPL1	001G516500.1	SbEXPL2	Bradi1g26990.2	BdEXPL1	AT3G45970	AtEXPL1
Os10g39640	OsEXPL2	GRMZM2G095968	ZmEXPL2	001G311000.1	SbEXPL1	Bradi1g28130.1	BdEXPL2	AT4G38400	AtEXPL2
Os07g29290	OsEXPL3	GRMZM2G026147	ZmEXPL3	002G300000.1	SbEXPL3	Bradi1g30050.1	BdEXPL3	AT3G45960	AtEXPL3
Os06g50960	OsEXPL4	GRMZM2G095968	ZmEXPL4	010G271000.1	SbEXPL4	Bradi1g76260.1	BdEXPL4	AT4G17030	AtEXPL4
Os07g31390	OsEXPL5	GRMZM2G004955	ZmEXPL5			Bradi1g76270.1	BdEXPL5		
						Bradi3g32297.1	BdEXPL6		

Fonte: Arquivo pessoal

APÊNDICE B

Nomeação dos transcritos de EXPA em cana-de-açúcar e identidade entre o grupo de sequências.

Nomeação	Representação	Código
ShEXPA15	1	SCCCLR1048B09.g
	2	R1_L38003_T1
	3	g110472_scaffold212368 size6181
	4	comp159274_c0_seq1
ShEXPA25	5	SCQSRT1034D12.g
	6	comp1383537_c0_seq1
ShEXPA27-zm	7	SCRURT2011A12.g
	8	g239175_scaffold656988 size280
ShEXPA16	9	lcl SCSP803280_000033064
	10	g266950_scaffold807208 size245
	11	SCJFRT1005B11.g
	12	g277422_scaffold870446 size233
ShEXPA17	13	lcl SCSP803280_000012816
	14	g35008_scaffold58740 size2090
	15	Locus45782.1
	16	comp1381849_c0_seq1
	17	SCRFL1010C03
ShEXPA19	18	lcl SCSP803280_000111101
	19	R1_L29364_T1
	20	g55853_scaffold109717 size2065
	21	lcl SP803280_c83446_g3_i1
	22	SCQGRT1040G03.g
ShEXPA27-sb	23	Locus3746.1
	24	SCVPRT2075H10.g
	25	comp197144_c0_seq1
	26	lcl SP803280_c103434_g2_i1
	27	R1_L12106_T2
	28	comp85689_c1_seq3
ShEXPA36	29	lcl SCSP803280_000095130
	30	g11878_scaffold18233 size3772
ShEXPA24	31	g145992_scaffold304081 size463
	32	R1_L28519_T1
ShEXPA8-sb	33	lcl SCSP803280_000039302
	34	lcl SP803280_c62089_g1_i1
ShEXPA8-zm	35	g91120_scaffold171349 size777
	36	g284546_scaffold916347 size225

Nomeação	Representação	Código
ShEXPA35	37	R1_L30976_T1
	38	SCCCFL2002C09.g
	39	lcl SP803280_c28875_g1_i1
ShEXPA1	40	SCJLRZ1019F12.g
	41	SCCCLB1024G11.g
	42	SCJFRZ2030D09.b
	43	comp59738_c0_seq1
	44	lcl SP803280_c104383_g3_i3
	45	comp79408_c0_seq2
ShEXPA14	46	g78032_scaffold147390 size982
	47	lcl SCSP803280_000043058
ShEXPA33	48	comp1098047_c0_seq1
	49	SCJLRT1018B08.g
	50	SCCCLR2C02A05.g
	51	R1_L10548_T1
	52	SCEQLB1064E10.g
	53	g140567_scaffold288782 size480
	54	SCSLV1004F11.g
	55	lcl SP803280_c104383_g3_i2
	56	Locus941.8_Length_1191
	57	Locus941.8_Length_1512
	58	Locus941.8_Length_1529
	59	Locus941.8_Length_1934
	60	Locus941.8_Length_846
	61	Locus941.8_Length_1473
	62	Locus941.8_Length_1517
	63	Locus941.8_Length_1451
	64	R1_L4895_T1
ShEXPA13	65	comp1465284_c0_seq1
	66	lcl SCSP803280_000190559
ShEXPA5	67	g44090_scaffold79858 size2081
	68	lcl SCSP803280_000096611
	69	lcl SP803280_c39784_g1_i1

Fonte: Arquivo pessoal

ShEXPA36				ShEXPA36											
29	30			29	30										
29	ID	0,985		29	ID	0,985									
30	0,99	ID		30	0,99	ID									
ShEXPA14				ShEXPA14											
47	48			47	48										
47	ID	0,99		47	ID	0,99									
48	0,99	ID		48	0,99	ID									
ShEXPA27-zm				ShEXPA27-zm											
7	8			7	8										
7	ID	0,95		7	ID	0,95									
8	0,95	ID		8	0,95	ID									
ShEXPA24				ShEXPA24											
31	32			31	32										
31	ID	0,99		31	ID	0,99									
32	0,99	ID		32	0,99	ID									
ShEXPA35				ShEXPA35											
37	38	39		37	38	39									
37	ID			37	ID										
38	1	ID		38	1	ID									
39	0,94	NA	ID	39	0,94	NA	ID								
ShEXPA8-sb				ShEXPA8-sb											
33	34			33	34										
33	ID	1		33	ID	1									
34	1	ID		34	1	ID									
ShEXPA13				ShEXPA13											
65	66			65	66										
65	ID	1		65	ID	1									
66	1	ID		66	1	ID									
ShEXPA5				ShEXPA5											
67	68			67	68										
67	ID	1		67	ID	1									
68	1	ID		68	1	ID									
ShEXPA8-zm				ShEXPA8-zm											
35	ZmEXPA8	36		35	ZmEXPA8	36									
35	ID	0,935	NA	35	ID	0,935	NA								
ZmEXPA8	0,94	ID	1	ZmEXPA8	0,94	ID	1								
36	NA	1	ID	36	NA	1	ID								
ShEXPA23															
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
49	ID														
50	1	ID													
51	1	1	ID												
52	1	1	1	ID											
53	0,93	0,926	0,93	0,93	ID										
54	1	1	1	1	0,93	ID									
55	1	1	1	1	0,93	1	ID								
56	1	1	1	1	0,93	1	1	ID							
57	1	1	1	1	0,93	1	1	1	ID						
58	0,95	0,951	0,95	0,95	0,9	0,95	0,95	0,95	0,95	ID					
59	1	1	1	1	0,93	1	1	0,61	0,61	0,56	ID				
60	1	1	1	1	0,93	1	1	1	1	0,95	0,6	ID			
61	1	1	1	1	0,93	1	1	1	1	0,95	0,61	1	ID		
62	1	1	1	1	0,93	1	1	1	1	0,95	0,61	1	1	ID	
63	1	1	1	1	0,93	1	1	1	1	0,95	0,56	1	1	1	ID
64	NA	NA	NA	1	0,91	0,98	1	1	1	0,96	1	1	1	1	ID

Fonte: Arquivo pessoal

Nomeação dos transcritos de EXPB em cana-de-açúcar e identidade entre o grupo de sequências.

Nomeação	Representação	Código
ShEXPB22	1	lcl SCSP803280_000113424
	2	g12122_scaffold18565 size3756
ShEXPB23	3	lcl SCSP803280_000108437
	4	lcl SP803280_c103899_g1_i1
	5	Locus448.10_Length_1077
	6	Locus448.10_Length_1067
	7	Locus448.10_Length_1108
	8	Locus448.10_Length_1046
	9	Locus448.10_Length_1105
	10	Locus448.10_Length_2412
	11	Locus448.10_Length_2332
	12	Locus448.10_Length_643
	13	SCQGR1045C07.g
	14	comp79247_c0_seq1
	15	R1_L42132_T1
ShEXPB2	16	comp73620_c0_seq1
	17	SCCST1004C07.g
	18	lcl SP803280_c53852_g1_i1
ShEXPB28	19	R1_L40358_T1
	20	R1_L11217_T1
	21	comp170010_c0_seq1
	22	lcl SP803280_c103899_g1_i2
	23	comp85849_c1_seq7
	24	SCVPRT2077B02.g
	25	Locus448.10_Length_783
	26	lcl SCSP803280_000065540
ShEXPB26	27	comp72913_c0_seq2_ok
	28	comp72334_c0_seq2_ok
	29	R1_L35282_T1
ShEXPB1	30	lcl SP803280_c104089_g1_i2
	31	lcl SCSP803280_000145538
	32	R1_L58758_T1
	33	comp195286_c1_seq1
	34	g37987_scaffold65652 size2087
	35	lcl SP803280_c106986_g1_i1
	36	comp83002_c0_seq2
ShEXPB4	37	R1_L32063_T1
	38	R1_L4110_T3
	39	lcl SP803280_c110956_g2_i2
	40	g75911_scaffold144125 size1031
	41	SCCCL4006H09.g
	42	comp1276431_c0_seq1
	43	lcl SCSP803280_000189163

Nomeação	Representação	Código
ShEXPB5	44	SCJLRT2050F08.g
	45	g64840_scaffold128820 size1500
ShEXPB24	46	g13858_scaffold20866 size3579
	47	lcl SCSP803280_000113567
ShEXPB31	48	lcl SCSP803280_000019559
	49	SCACAD1037F10.g
	50	lcl SP803280_c76612_g1_i1
	51	g1684_scaffold1655 size6042
ShEXPB30	52	lcl SCSP803280_000098880
	53	comp184933_c0_seq1
	54	R1_L55763_T1
	55	lcl SP803280_c82524_g1_i1
	56	R1_L59293_T1
ShEXPB3	57	SCJLLR1033E07.g
	58	SCQGHR1010G04.g
	59	comp82804_c1_seq2
ShEXPB9	60	lcl SCSP803280_000155728
	61	lcl SP803280_c110956_g2_i4
	62	R1_L13596_T1
ShEXPB12	63	SCCCLR1065A10.g
	64	R1_L4645_T1
	65	comp72968_c0_seq1
ShEXPB21	66	SCACLR1036F02.g
	67	SCMCRT2107D08.g
	68	comp68491_c0_seq1
	69	lcl SP803280_c94617_g2_i1
	70	comp60230_c1_seq1
ShEXPB25	71	SCEZR21012B06.g
	72	comp79768_c0_seq2
	73	Locus32504.3_Length_449
	74	lcl SP803280_c92149_g2_i1
	75	SCCCHR1002H02.g
	76	comp76715_c1_seq1
	77	Locus32504.3_Length_457
	78	lcl SCSP803280_000040604
ShEXPB29	79	lcl SCSP803280_000147924
	80	SCCCLB2007H07.g
ShEXPB33	81	lcl SCSP803280_000064994
	82	g23202_scaffold32002 size2237
	83	SCAGSD1041B02.g
ShEXPB6	84	SCCCL3120E05.g
	85	R1_L7209_T1
	86	comp82724_c2_seq3

Fonte: Arquivo pessoal

ShEXPB2

16	17	18	19
16	ID		
17	1	ID	
18	1	1	ID
19	1	1	1 ID

ShEXPB3

57	58	59
57	ID	
58	0,99	ID
59	1	1 ID

ShEXPB5

44	45
44	ID 0,9
45	0,94 ID

ShEXPB6

84	85	86
84	ID 1	1
85	1 ID	1
86	1	1 ID

ShEXPB9

60	61	62
60	ID	
61	1 ID	
62	1	1 ID

ShEXPB25

71	72	73	74	75	76	77	78
71	ID						
72	1	ID					
73	1	1	ID				
74	1	1	1	ID			
75	0,96	NA	NA	0,96	ID		
76	1	NA	NA	1	0,96	ID	
77	1	1	1	1	NA	NA	ID
78	1	1	1	1	0,96	1	1 ID

ShEXPB28

20	21	22	23	24	25	26
20	ID					
21	1	ID				
22	1	1	ID			
23	1	0,9	0,96	ID		
24	1	1	1	0,96	ID	
25	1	1	0,99	0,95	0,99	ID
26	1	1	1	0,96	1	0,99 ID

ShEXPB1

31	32	33	34	35	36	37
31	ID	1	0,94	1	1	0,97 1
32	1	ID	0,97	0,97	0,97	1 NA
33	0,9	1	ID	0,94	0,94	0,97 0,97
34	1	1	0,94	ID	1	0,97 NA
35	1	1	0,94	1	ID	0,97 1
36	1	1	0,97	0,97	0,97	ID 1
37	1	NA	0,97	NA	1	1 ID

ShEXPB4

38	39	40	41	42	43
38	ID				
39	0,93	ID			
40	0,93	1	ID		
41	0,93	1	1	ID	
42	0,84	1	1	1	ID
43	1	1	1	1	1 ID

ShEXPB30

52	53	54	55	56
52	ID			
53	0,98	ID		
54	0,98	1	ID	
55	1	1	0,98	ID
56	0,85	NA	NA	0,85 ID

ShEXPB31

48	49	50	51
48	ID		
49	1	ID	
50	1	1	ID
51	0,94	0,6	0,94 ID

ShEXPB21

66	67	68	69	70
66	ID			
67	1	ID		
68	1	1	ID	
69	1	NA	NA	ID
70	0,91	NA	NA	0,91 ID

ShEXPB12

63	64	65
63	ID 1	1
64	1	ID 1
65	1	1 ID

ShEXPB22

1	2
1	ID 1
2	1 ID

ShEXPB24

46	47
46	ID 1
47	0,97 ID

ShEXPB29

79	80
79	ID 0,99
80	0,99 ID

ShEXPB26

27	28	29	30
27	ID		
28	1	ID	
29	1	1	ID
30	1	1	1 ID

ShEXPB33

81	82	83
81	ID 0,95	0,9
82	0,95	ID 0,95
83	0,9	0,95 ID

ShEXPB23

3	4	5	6	7	8	9	10	11	12	13	14	15
3	ID											
4	1	ID										
5	0,99	0,99	ID									
6	0,99	0,99	1	ID								
7	0,99	0,99	1	1	ID							
8	0,99	0,99	1	1	1	ID						
9	0,99	0,99	1	1	1	1	ID					
10	0,99	0,99	1	1	1	1	1	ID				
11	0,99	0,99	1	1	1	1	1	1	ID			
12	0,96	0,96	0,97	0,97	0,97	0,97	0,97	0,97	0,97	ID		
13	0,99	0,99	1	1	1	1	1	1	1	1	ID	
14	1	1	1	1	1	1	0,65	0,65	1	1	NA	NA ID
15	1	1	1	1	1	1	1	1	1	0,92	NA	NA ID

Fonte: Arquivo pessoal

Nomeação dos transcritos de EXPL em cana-de-açúcar e identidade entre o grupo de sequências.

Nomeação	Representação	Código
ShEXPL1	1	R1_L12741_T1
	2	SCSBSD2058D09.g
	3	comp80035_c1_seq3
	4	SP803280c98251g3i1
ShEXPL2	5	SP803280c114720g1i1
	6	lcl SCSP803280_000124172
	7	comp181173_c2_seq1
	8	SCJFRZ2014H02.g
	9	g112570_scaffold217259 size605
	10	comp88313_c0_seq5
ShEXPL3	11	comp80035_c1_seq4
	12	SP803280c95964g3i1
	13	Locus15875.3_Confidence_0.636_Length_1444
	14	SCAGLR1064B01.g
	15	Locus15875.3_Confidence_0.636_Length_1456
	16	SCJFLR2036G07.g
	17	comp196627_c0_seq1
	18	comp73362_c0_seq4
ShEXPL4	19	SCSGHR1067F05.g
	20	Locus45283.1_Confidence_1.000_Length_224
	21	lcl SP803280_c90616_g1_i1
	22	comp538210_c0_seq1
	23	comp174239_c0_seq1

ShEXPL1

	1	2	3	4
1	ID			
2	0,99	ID		
3	1	0,99	ID	
4	0,99	0,98	1	ID

ShEXPL2

	5	6	7	8	9	10
5	ID					
6	1	ID				
7	1	1	ID			
8	1	1	1	ID		
9	0,98	0,977	NA	0,98	ID	
10	1	1	NA	1	NA	ID

ShEXPL3

	11	12	13	14	15	16	17	18
11	ID							
12	0,99	ID						
13	0,99	0,992	ID					
14	0,99	0,978	1	ID				
15	0,99	0,992	1	0,99	ID			
16	1	1	NA	NA	NA	ID		
17	1	1	NA	NA	NA	1	ID	
18	NA	NA	1	1	1	NA	NA	ID

ShEXPL4

	19	20	21	22	23
19	ID				
20	0,9	ID			
21	0,9	1	ID		
22	1	1	1	ID	
23	NA	NA	1	NA	ID

Fonte: Arquivo pessoal

APÊNDICE C

Alinhamento múltiplo das sequências de EXPA de sorgo com sequências de EXPA e EXPB de tomate. Destacado em vermelho estão os motivos conservados, linhas indicam a localização do peptídeo final, α e β -inserção e setas indicam o fim e o começo dos domínios D1 e D2 respectivamente.

Fonte: Arquivo pessoal

Alinhamento múltiplo das sequências de EXPB de sorgo com sequências de EXPA e EXPB de tomate. Destacado em vermelho estão os motivos conservados, linhas indicam a localização do peptídeo final, α e β -inserção e setas indicam o fim e o começo dos domínios D1 e D2 respectivamente.

Signal peptide

Fonte: Arquivo pessoal

Alinhamento múltiplo das sequências de EXPL de sorgo com sequências de EXPL de tomate. As setas indicam o fim e o começo dos domínios D1 e D2, respectivamente.

Fonte: Arquivo pessoal

APÊNDICE E

Motivos conservados da subfamília EXPA de sorgo encontrados pelo programa MEME (<http://meme-suite.org/tools/meme>)

Motivo 1

Motivo 2

Motivo 3

Motivo 4

Motivo 5

Motivo 6

Motivo 7

Motivo 8

Motivo 9

Motivo 10

Fonte: Arquivo pessoal

Motivos conservados da subfamília EXPB de sorgo encontrados pelo programa MEME (<http://meme-suite.org/tools/meme>)

Fonte: Arquivo pessoal

Motivos conservados da subfamília EXPL de sorgo encontrados pelo programa MEME (<http://meme-suite.org/tools/meme>)

Fonte: Arquivo pessoal

APÊNDICE F

Análise de éxons e íntrons da subfamília EXPA em sorgo.

Fonte: Arquivo pessoal

Análise de éxons e íntrons da subfamília EXPB em sorgo.

Fonte: Arquivo pessoal

Análise de éxons e íntrons da subfamília EXPL em sorgo.

Fonte: Arquivo pessoal

APÊNDICE G

Desenho dos iniciadores das sequências de cana-de-açúcar obtidas a partir dos diferentes bancos de transcritos em EXPA e EXPB.

Nome		Sequência 5' - 3'
ShEXPA1	F	GCTTTGGACAGACGTTACC
	R	CTCAACTGGGCTAGCAGCT
ShEXPA4	F	TCATTCCATCCACAAGCTCA
	R	TCCAGATCCCTGATCCTGTG
ShEXPA5	F	GGTGGGCAGTACATCCAGTT
	R	ATGCAAGCAGCAGAGAGTGA
ShEXPA8	F	CAGTTCGGCCAGACCTACTC
	R	AGGAAAAGGAGCGGATTGT
ShEXPA9	F	TAAATACACTCGCCCTTGC
	R	ACAGCCTTCTTGGCACTTGT
ShEXPA12	F	GGTGCAGATCAAAGGCTCAT
	R	AGCATCTCCTCCGTGTCAC
ShEXPA13	F	AAACATACCAGGCCACCAAG
	R	ATTGACTGCATGCATGGTTT
ShEXPA14	F	AGTTTGGCAGCACCTACCAG
	R	TACGACACCTGCAGCAGAAC
ShEXPA15	F	CTTCGGCCAGACCTTCAG
	R	TCGTGGTAGAGTCGTCTCC
ShEXPA16	F	GACCTACACCGGCAACAGT
	R	GAGGAGGGCTCTGTTTGAA
ShEXPA17	F	GCTTTGGACAGACGTTACC
	R	CTCAACTGGGCTAGCAGCT
ShEXPA19	F	CCAGCTCTGTGAGGTGAGGT
	R	GTTCTTGATTCTGTCTGGT
ShEXPA24	F	AGAGCAACGCGTACCTCAAC
	R	CACCTACACGTCCAATGCTC
ShEXPA25	F	GCTCCAGCTCGAGATTTCAT
	R	ACATGCCACCTCTACGGAG
ShEXPA27-sb	F	CAAGACGCTGACCAGTTTCA
	R	AGAATGAGTGCCGCAAGATC
ShEXPA27-zm	F	GGTACGGCAACCTGTAICTG
	R	CAGTGCTACAAGATCGCGTG
ShEXPA28	F	TTTGGCAATGATCTGAACCA
	R	TAGTGCGTCGTCTGCTCT
ShEXPA31	F	AACAACGGGTCGTCTCTC
	R	TCAACGGCTTCTACATGGTG
ShEXPA32	F	GTACCAGGCGTCCAAGAAGT
	R	CTCAATACGGTGCCAGAGT
ShEXPA33	F	GGCAGCTGACGATATAGCAA
	R	GTGCTCTTTCTCTCTGCTC
ShEXPA34	F	CTCCTACCAGGGATCCAACA
	R	GTTGTTTTTCGCTTCTGGA
ShEXPA35	F	ACGTCAAGGAGTTCAGTCG
	R	CTTCAGGTACGGCCGGTAC
ShEXPA36	F	AGAGCAGGCTGCAGTTCAA
	R	GTGCTTTGCTGGTGGATCTT

Nome		Sequência 5' - 3'
ShEXPB1	F	GTCACCTCGTCGTGTTGCTA
	R	ATTTTGGTGGTCTGTGAGC
ShEXPB2	F	ACTGTTGTCATCACCGACCA
	R	TGACATGAGTGGGACTGCAT
ShEXPB3	F	GCTCCTTCTGTCCAGTACCAG
	R	ATTGGAGTGGAGCTGGAGCT
ShEXPB4	F	GCGGCCAACACCTATTACC
	R	TCAACATTGGCTTTCATTGG
ShEXPB5	F	ACGAGCCACTGTTCAAGGAC
	R	TGAACACTATCCGGTCGCC
ShEXPB6	F	ACTACCGTTCCATCGTCCAG
	R	GTATTGTGTGTCGATGGG
ShEXPB9	F	GAAGCCATTGACCGACTACC
	R	GGCTACAACCTCGTTCCATT
ShEXPB10	F	AGTACAGGAGGTCCGTGGTG
	R	TCTACTTTCGGTGATCGTG
ShEXPB11	F	TAGGAGCGGTAGTCGGTGTT
	R	TCTCCTTCAAGCCTTTGC
ShEXPB12	F	GATTACCGCTCCTACGTCCA
	R	CAACCGGACTGAGGATTTTC
ShEXPB13	F	AAGAAGCTCATCGCCAAAGA
	R	TTTTTGATGCCCTTCGTCC
ShEXPB17	F	AAGACTGGAAGCCCGACAC
	R	TTACCATACCTGGGTTGGCT
ShEXPB18	F	GAGGCACCAAGAGCGTCTAC
	R	AGACCTACTGCTCTGCGG
ShEXPB19	F	CATCATCGACATGCAGTTCC
	R	CTGCTGGTCAAGTACGTCGA
ShEXPB21	F	ACAGCTCCATCGTCCAGTTC
	R	TCTTCAATGCAATGCAAGC
ShEXPB22	F	ACTGGAGGCCAAACACATTC
	R	GGATTGCCATTGCCATTATT
ShEXPB23	F	GCAAGACCTACCGTTCCTTG
	R	GTGCATAAATTAAGCGGGGA
ShEXPB24	F	GACCTACCGTTCTTGGTCA
	R	GGTGTACAGGTGAGAAGAGG
ShEXPB25	F	TGGACTGCACACGGAAATTA
	R	GCTGTAGGTGGTGTCTGTG
ShEXPB26	F	TACACTTCCCGGCTCAATTT
	R	GTTGTAAGATGGCAACGGG
ShEXPB27	F	CCGCTCCTTGGTGAAGTACT
	R	TATGCGTGGTCAAGAAGTGC
ShEXPB28	F	CCGCTCCTTGGTGAAGTACT
	R	GAGTTCTTCCCTCCCTCCAC
ShEXPB29	F	CTCGGTGCTCAACTTCAACA
	R	GGATGGAGTGGATCGACG
ShEXPB30	F	GACGTCATCCCAGGAAGT
	R	TAGGAGAGCCAGAGACTGCC
ShEXPB31	F	AAGTCTCGCCCTCTCAACAA
	R	CATCCGAGGTCCATGTCTCT
ShEXPB33	F	TCGTCAACTTCGGAACAAAC
	R	AGAACCTACCTCTGTCCCGG

Fonte: Arquivo pessoal

Desenho dos iniciadores das sequências de cana-de-açúcar obtidas a partir dos diferentes bancos de transcritos em EXPL.

Nome		Sequência 5' - 3'
ShEXPL1	F	GAGGATACCCTGCGAGTACG
	R	CCTCTCCATCAGGTTCTGT
ShEXPL2	F	CGTCCAGATCACTGACATCG
	R	AAGCAGAAGCTGGTCCTGAA
ShEXPL3	F	CGAGATGGACGACTGGAAG
	R	GACAGCATGCAATGCAAATG
ShEXPL4	F	ACAAGAACCTGTCGGTACGC
	R	ACATCCCTGTCGTCGAGATC

Fonte: Arquivo pessoal