

**VANIA GABRIELA SEDANO PARTIDA**

**Caracterização genômica dos genes *S6PP* de *Saccharum*  
*spp.* associados no metabolismo da sacarose.**

Dissertação apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia da Universidade de São Paulo, Instituto Butantã e Instituto de Pesquisas Tecnológicas para obtenção do Título de Mestre em Biotecnologia.

São Paulo

2018

**VANIA GABRIELA SEDANO PARTIDA**

**Caracterização genômica dos genes *S6PP* de *Saccharum* spp.  
associados no metabolismo da sacarose.**

Dissertação apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia da Universidade de São Paulo, Instituto Butantã e Instituto de Pesquisas Tecnológicas para obtenção do Título de Mestre em Biotecnologia.

Área de Concentração: Biotecnologia

Orientador (a); Profa. Dra. Marie-Anne Van Sluys

Versão corrigida. A versão original eletrônica encontra-se disponível tanto na Biblioteca do ICB quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

São Paulo

2018

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
Serviço de Biblioteca e informação Biomédica  
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Sedano Partida, Vania Gabriela  
Caracterização genômica dos genes S6PP de  
Saccharum spp. associados no metabolismo da  
sacarose. / Vania Gabriela Sedano Partida;  
orientador Marie-Anne Van Sluys. -- São Paulo, 2018.  
81 p.

Dissertação (Mestrado) ) -- Universidade de São  
Paulo, Instituto de Ciências Biomédicas.

1. Sucrose. 2. S6PP. 3. Cana-de-açúcar. 4.  
Expressão diferencial. 5. Proteína recombinante. I.  
Van Sluys, Marie-Anne, orientador. II. Título.

UNIVERSIDADE DE SÃO PAULO  
**Programa de Pós-Graduação Interunidades em Biotecnologia**  
Universidade de São Paulo, Instituto Butantan, Instituto de Pesquisas Tecnológicas

Candidato(a): Vania Gabriela Sedano Partida

Titulo da Dissertação: Caracterização genômica dos genes *S6PP* de *Saccharum* spp. associados no metabolismo da sacarose.

Orientador: Marie-Anne Van Sluys

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado, em sessão publica realizada a ...../...../....., considerou o(a) candidato(a):

**Aprovado(a)**       **Reprovado(a)**

Examinador(a):                      Assinatura: .....

Nome: .....

Instituição: .....

Examinador(a):                      Assinatura: .....

Nome: .....

Instituição: .....

Examinador(a):                      Assinatura: .....

Nome: .....

Instituição: .....

Presidente:                              Assinatura: .....

Nome: .....

Instituição: .....



UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"  
Av. Prof. Lineu Prestes, 2415 - CEP. 05508-000 São Paulo, SP - Brasil  
Telefone : (55) (11) 3091-7733 - telefax : (55) (11) 3091-8405  
e-mail: [cep@icb.usp.br](mailto:cep@icb.usp.br)

*Comissão de Ética em Pesquisa*

## CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB Nº **723/15** referente ao projeto intitulado: "*Caracterização genômica dos genes *s6pp* de *Saccharum spp.* vinculados na produção de sacarose*" sob a responsabilidade de **Vânia Gabriela Sedano Partida**, foi analisado na presente data pela CEUA - COMISSÃO DE ÉTICA NO USO DE ANIMAIS e pela CEPSh- COMISSÃO DE ÉTICA EM PESQUISA COM SERES HUMANOS, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da lei 11.794 de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº466 de 2012.

São Paulo, 24 de março de 2015.

PROF. DR. WOTHAN TAVARES DE LIMA  
Coordenador da CEUA - ICB/USP

PROF. DR. PAOLO M.A ZANOTTO  
Coordenador da CEPsh - ICB/USP

*A Sebastián, por supuesto.*

## AGRADECIMENTOS

Agradezco a mis padres por siempre ir junto conmigo en este camino, por apoyarme y nunca soltarme de la mano aún en la distancia. A mi hermana por ser uno de mis más grandes pilares en este recorrido y mi mejor ejemplo de dedicación. Al gran amor que Brasil me dio, mi esposo Gustavo, por nunca dudar de mí, por apoyarme siempre, por darle alegría a mis días y hacer más ameno mi andar. A mi pequeño Sebastián, que con su llegada cambió mi vida y aunque me hizo los días más cortos y más cansados, también me ha traído las más grandes satisfacciones que nunca creí que podría sentir. Simplemente se volvió mi fuente de inspiración.

A mi familia en general, que desde lejos siento su apoyo, a mis primos, mis tíos, mis abuelos, porque la distancia ha creado vínculos más fuertes entre nosotros. Y a todos los seres que han formado parte de estos tres años, Salomón y Mateo, por traerme tanta felicidad.

Agradezco firmemente a la profesora Marie-Anne Vans Sluys, mi orientadora, por abrirme las puertas de este laboratorio, por darme la oportunidad de ser parte de este equipo, por transmitirme tanta experiencia y conocimiento y por brindarme su apoyo siempre que lo necesité, porque más que profesora, fue amiga, siempre mostrando interés por mí y mi familia.

A todo el equipo del GaTE Lab, que sin duda fueron mi segunda familia. Agradezco a quienes me recibieron la primera vez que llegué a este laboratorio, Sarah Oliveira, Dani Quintanilla y Erika Maria de Jesús, porque gracias a ustedes es que tomé la decisión de emprender este viaje. A Paula Turrini, Andréia Prata, Gesiele Barros, Jonas Gaiarsa, Andrés Ochoa, Augusto Crivellari, Juliane Ishida y Henrique Dias que siempre estuvieron a disposición y de alguna manera contribuyeron en el conocimiento que adquirí durante estos tres años. Un agradecimiento muy especial a Tati por endulzarnos los días con esos bolos tan deliciosos!!! Y por ser siempre muy profesional en su trabajo y aunque me llevé unas jaladas de oreja, siempre me demostró que puedo contar con ella para lo que sea. A la loira Raquel Miranda y José Hernandez por ser mis compañeros y amigos, por hacerme reír a carcajadas con sus ocurrencias y peleas, haciendo los días más divertidos. A Geovani Tolfo que siempre estuvo de mi lado como un hermano, dando soporte en cualquier cosa y especialmente por ayudarme en el examen de admisión.

Al personal del laboratório de Proteínas e Biologia Redox del professor Dr. Luis Eduardo Soares Netto por todo el apoyo brindado siempre, en especial a Simone, Thiago, Fernando, Angélica y sobretudo a Renata Bannitz por el conocimiento que me transmitió, por siempre estar a mi lado en el momento que la necesité y por brindarme su amistad.

Al laboratorio de Biologia Estrutural e Funcional de Sistemas de Secreção de Macromoléculas e Vias de Sinalização em Bactérias del profesor Dr. Shaker Chuck Farah, por apoyarnos con equipo de laboratorio para desenvolver el proyecto, en especial a Edgar Llontop quien me recepcionó y ayudó siempre que lo solicité.

A las agencias de fomento que nos dieron el medio financiero para realizar este proyecto, a CAPES por otorgarme la beca durante el periodo comprendido entre 2015 al 2017, referente al numero de proceso 1523830 y al proyecto temático FAPESP 2008/52074-0.

**“Quem nunca errou, nunca experimentou  
nada novo”.**

**(Albert Einstein)**

## RESUMO

SEDANO PARTIDA, Vania Gabriela. **Caracterização genômica dos genes *S6PP* de *Saccharum sp.* associados no metabolismo da sacarose.** 2018. 81 f. Dissertação (Mestrado em Interunidades em Biotecnologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2018.

A sacarose é um tipo de açúcar formado durante a fotossíntese das plantas e é armazenada no vacúolo das células dos tecidos não fotossintéticos, fornecendo uma fonte de energia essencial para realizar processos fundamentais no crescimento da planta. A sacarose é principalmente extraída da cana-de-açúcar, devido à sua capacidade de armazenar grandes quantidades de sacarose nos entrenós dos seus colmos, de onde é extraída. A cana-de-açúcar é considerada uma fonte de alimento e de energia renovável e o seu cultivo resulta em um aumento na economia convertendo a indústria sucroalcooleira como um dos setores mais modernos e competitivos no mundo. Brasil tem sido líder mundial em produção durante vários anos consecutivos e devido à grande importância econômica, este projeto visa ao entendimento genético dos genes que dão lugar à expressão da enzima, sacarose 6-fosfato fosfohidrolase (*S6PP*), responsável pela liberação da sacarose no último passo da síntese. Pretende-se entender a função de dois possíveis genes homólogos *S6PP* identificados no genoma da cultivar híbrida de *Saccharum spp.*, variedade R570, onde verificou-se a existência de uma duplicação gênica em tandem (*S6PP.1* e *S6PP.2*) compartilhada com outros genomas de plantas, que resulta na predição de transcrição de dois peptídeos iguais. A relevância que introduzimos neste trabalho é a proposta de uma nova isoforma desse gene nunca evidenciada em outros trabalhos, que sugere a conjugação destas duas sequências em tandem, evento possivelmente decorrente de *splicing* alternativo, resultando em um único transcrito (*S6PP\_2D*) contendo os dois domínios da *S6PP*, surgindo a hipótese de que *Saccharum spp.* apresenta duas variantes do gene *S6PP*. É por isso que este trabalho compreende o estudo molecular dos genes *S6PP* em diferentes tecidos de variedades de cana-de-açúcar ao longo do tempo, onde foi evidenciada a expressão do transcrito da versão inédita *S6PP\_2D* proposta neste trabalho, sugerindo que atua na liberação da sacarose durante a síntese e durante o armazenamento. Foi estabelecida a função bioquímica das proteínas sintetizadas em organismos heterólogos.

**Palavras-chave:** Sacarose. *S6PP*. Cana-de-açúcar. Expressão diferencial. Proteína recombinante.

## ABSTRACT

SEDANO PARTIDA, Vania Gabriela. **Genomic characterization of *S6PP* genes from *Saccharum* sp. associated in sucrose metabolism.** 2018. 81 f. Dissertation (Master Interunits in Biotechnology) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2018.

Sucrose is a type of sugar formed during plants photosynthesis and is stored in the vacuole of the cells of non-photosynthetic tissues, providing a source of essential energy to perform fundamental processes in plant growth. Sucrose is mainly extracted from sugarcane, due to its ability to store large amounts of sucrose in the internodes of its stems, from where it is extracted. Sugarcane is considered a source of food and renewable energy and its cultivation results in an increase in the economy, making the sugar and alcohol industry one of the most modern and competitive sectors in the world. Brazil has been a world leader in production for several consecutive years and because of the great economic importance, this project aims at the molecular understanding of genes that give rise to the expression of the enzyme sucrose 6-phosphate phosphohydrolase (*S6PP*), responsible for the release of sucrose in the last synthesis step. It is intended to understand the function of two possible homologous *S6PP* genes identified in the genome of the hybrid cultivar *Saccharum* spp., Variety R570, where there was a tandem gene duplication (*S6PP.1* and *S6PP.2*) shared with others genomes of plants, which results in the prediction of transcription of two equal peptides. The relevance that we introduce in this work is the proposal of a new isoform of this gene never evidenced in other works, that suggests the conjugation of these two sequences in tandem, event possibly resulting from alternative splicing, resulting in a single transcript (*S6PP\_2D*) containing the two domains of the *S6PP*, with the hypothesis that *Saccharum* spp. presents two variants of the *S6PP* gene. This is why this work comprises the molecular study of the *S6PP* genes in different tissues of sugarcane varieties over time, where it was evidenced the expression of the transcript of the unpublished version *S6PP\_2D* proposed in this work, suggesting that it acts in the release of sucrose during synthesis and during storage. The biochemical function of proteins synthesized in heterologous organisms has been established.

**Keywords:** Sucrose. *S6PP*. Sugarcane. Differential expression. Recombinant protein.

## LISTA DE FIGURAS

Figura 1.	Metabolismo geral da sacarose.....	20
Figura 2.	Últimos passos da via metabólica da síntese de sacarose.....	22
Figura 3.	Quantidades de produção em toneladas de cana-de-açúcar por país.....	27
Figura 4.	Reação catalisada pela sacarose 6-fosfato fosfohidrolase (S6PP).....	29
Figura 5.	Representação gráfica das isoformas dos genes <i>S6PP</i> , encontradas em dois diferentes fragmentos genômicos.....	32
Figura 6.	Desenho de primers sobre os CDSs preditos dos genes <i>S6PP.1</i> e <i>S6PP_2D</i> envolvendo a junção exon-exon .....	41
Figura 7.	Representação gráfica dos fragmentos genômicos BACs que contém as sequências dos genes <i>S6PP</i> .....	51
Figura 8.	Filogenia da região em tandem <i>S6PP</i> de BACs de cana-de-açúcar híbridas, R570 e SP80-3280 .....	52
Figura 9.	Árvore filogenética dos domínios protéicos da proteína S6PP de vários genomas de plantas.....	55
Figura 10.	Gráfico de barras dos valores de FPKM determinado a partir da análise da expressão por RNAseq de amostras de cana-de-açúcar R570.....	57
Figura 11.	Gráfico de barras dos valores de FPKM determinado a partir da análise da expressão por RNAseq de amostras de cana-de-açúcar SP70-1143.....	58
Figura 12.	Gráfico de barras dos valores de RPKM determinado a partir da análise da expressão por RNAseq de amostras de cana-de-açúcar RB92-5345.....	59
Figura 13.	Gráfico de barras do teste de genes normalizadores para qPCR.....	61

Figura 14.	Gráfico de barras da análise de expressão dos genes <i>S6PP</i> ao longo do tempo (3, 6 e 9 meses) em folha, meristema, colmo e raiz.....	62
Figura 15.	Modelo do ciclo da sacarose durante o desenvolvimento da planta em idades jovens (planta imatura) e idades velhas (planta madura).....	64
Figura 16.	Mapa biológico da expressão dos genes <i>S6PP.1</i> e <i>S6PP_2D</i> nos diferentes tecidos analisados.....	66
Figura 17.	Cromatograma da purificação da proteína <i>S6PP.1</i> .....	68
Figura 18.	SDS-PAGE dos conjuntos das frações da purificação da proteína <i>S6PP.1_pET15b</i> concentradas com filtro de 30 kDa e 100 kDa.....	69
Figura 19.	Cromatograma da purificação da proteína <i>S6PP_2D</i> .....	70
Figura 20.	SDS-PAGE do conjunto de frações concentradas provenientes da purificação da proteína <i>S6PP_2D</i> .....	71
Figura 21.	Atividade enzimática das proteínas heterólogas <i>S6PP.1</i> e <i>S6PP_2D</i> expressadas no vetor pET15b.....	72
Figura 22.	Gráficos de linha do ensaio enzimático com variação de tempo e temperatura.....	73
Figura 23.	Ensaio enzimático da proteína <i>S6PP.1</i> pura e com extratos enzimáticos semi-puros.....	74

## LISTA DE TABELAS

Tabela 1.	Resumo de estudos da enzima S6PP realizados em plantas e bactérias.....	30
Tabela 2.	Lista dos fragmentos genômicos que contém os genes <i>S6PP</i> .....	36
Tabela 3.	Lista de primers utilizados nos experimentos de PCR e qPCR.....	40
Tabela 4.	Amostras utilizadas para análise da expressão por qPCR.....	42
Tabela 5.	Variáveis testadas na expressão da proteína recombinante.....	46
Tabela 6.	Lista dos fragmentos genômicos identificados em 21 BACs contendo a duplicação dos genes <i>S6PP</i> .....	50
Tabela 7.	Lista de plantas que contém a duplicação em tandem dos genes <i>S6PP</i> no seu genoma (phytozome12).....	53
Tabela 8.	Bancos de dados de RNAseq gerado a partir do sequenciamento de diferentes bibliotecas de cana-de-açúcar sometidas em diferentes condições de desenvolvimento.....	56
Tabela 9.	Variáveis selecionadas após a padronização da expressão, para indução em larga escala e método de purificação.....	67

## SUMÁRIO

1. INTRODUÇÃO .....	17
1.1. A sacarose .....	18
1.2. Metabolismo da sacarose .....	19
1.2.1. Fotossíntese em cana-de-açúcar; metabolismo C4.....	20
1.2.2. Síntese da sacarose em plantas.....	22
1.2.3. Transporte da sacarose através do floema.....	23
1.2.4. Armazenamento da sacarose.....	23
1.3. Cana-de-açúcar.....	24
1.3.1. Cultivares híbridos R570 e SP80-3280 .....	25
1.3.2. A cana-de-açúcar no mundo.....	26
1.3.3. Distribuição dos cultivos de cana-de-açúcar no Brasil.....	27
1.3.4. Importância econômica.....	28
1.4. Enzima sacarose 6-fosfato-fosfohidrolase (S6PP).....	29
1.5. Estudos prévios dos genes S6PP no Genomics and Transposable Elements Laboratory (GaTE Lab).....	31
2. HIPÓTESE.....	33
3. OBJETIVO.....	34
4. METODOLOGIA.....	35
4.1. ESTUDO MOLECULAR DO ENTORNO GENÔMICO DOS GENES S6PP.....	35
4.1.1. Genômica comparativa.....	35
4.1.2. Análise de expressão em bancos de dados públicos de RNAseq.....	36
4.1.3. Perfil da expressão dos genes S6PP em tecidos de cana-de-açúcar.....	38
4.1.3.1. <i>Material vegetal e condições de cultivo</i> .....	38
4.1.3.2. <i>Extração e tratamento do RNA total</i> .....	38
4.1.3.3. <i>Transcrição reversa para síntese de cDNA</i> .....	39
4.1.3.4. <i>Análise da qualidade da reação de síntese de cDNA</i> .....	39
4.1.3.5. <i>Busca do melhor gene normalizador para quantificação da</i>	

<i>expressão por qPCR</i> .....	40
4.1.3.6. <i>Desenho dos iniciadores para quantificação dos genes pela técnica de qPCR</i> .....	41
4.1.3.7. <i>Análise quantitativo da expressão dos genes S6PP por qPCR</i> .....	42
<b>4.2. ESTUDO BIOQUÍMICO E FUNCIONAL DAS PROTEÍNAS S6PP EXPRESSADAS EM SISTEMAS HETERÓLOGOS</b> .....	<b>42</b>
4.2.1. <i>Síntese das proteínas S6PP em linhagens de expressão de E. coli</i> .....	42
4.2.1.1. <i>Preparo de DNA plasmidial em pequena escala (Miniprep)</i> .....	42
4.2.1.2. <i>Construção gênica dos genes S6PP dentro do plasmídeo de expressão pET15b</i> .....	43
4.2.1.3. <i>Construção gênica dos genes S6PP dentro do plasmídeo de expressão pETSUMO</i> .....	45
4.2.1.4. <i>Padronização da expressão da proteína recombinante</i> .....	46
4.2.1.5. <i>Testes de solubilidade para eleição da melhor cepa de expressão</i> ....	47
4.2.1.6. <i>Indução em larga escala para purificação da proteína recombinante</i> .....	47
4.2.1.7. <i>Padronização da purificação da proteína recombinante pelo método de IMAC</i> .....	48
4.2.2. <i>Ensaio de atividade enzimática das duas versões de S6PP</i> .....	49
<b>5. RESULTADOS E DISCUSSÃO</b> .....	<b>50</b>
5.1. <i>ESTUDO MOLECULAR DO ENTORNO GENOMICO DOS GENES S6PP</i> .....	50
5.1.1. <i>Genômica comparativa</i> .....	50
5.1.2. <i>Análises filogenéticas em plantas</i> .....	53
5.1.3. <i>Análise de expressão em bancos de dados públicos de RNAseq</i> .....	55
5.1.3.1. <i>Meristema e raiz</i> .....	57
5.1.3.2. <i>Inoculação com Sporisorium scitamineum</i> .....	58
5.1.3.3. <i>Inoculação com Acidovorax avenae subsp. avenae</i> .....	59
5.1.4. <i>Perfil da expressão dos genes S6PP em diferentes tecidos de cana-de-açúcar ao longo do desenvolvimento</i> .....	60
5.1.4.1. <i>Busca do melhor gene normalizador para quantificação da</i>	

<i>expressão por qPCR.....</i>	60
5.1.4.2. <i>Análise quantitativo da expressão dos genes S6PP por qPCR.....</i>	61
5.2. ESTUDO BIOQUIMICO E FUNCIONAL DAS PROTEÍNAS S6PP EXPRESSADAS EM SISTEMAS HETERÓLOGOS.....	67
5.2.1. Padronização da expressão da proteína recombinante e método de purificação.....	67
5.2.2. Indução em larga escala e purificação da proteína recombinante.....	67
5.2.3. Purificação da enzima S6PP.1.....	68
5.2.4. Purificação da enzima S6PP_2D.....	70
5.2.5. Ensaio de atividade enzimática das versões de S6PP.....	71
5.2.6. Ensaio enzimático com extratos proteicos semi-puros.....	72
5.2.7. Ensaio enzimático com extratos proteicos semi-puros com variação de tempo e temperatura de reação.....	72
5.2.8. Ensaio enzimático com a enzima S6PP.1 pura e com frações semi-puras.....	74
6. CONCLUSÕES.....	75
REFERENCIAS.....	77

## 1 INTRODUÇÃO

A sacarose é a forma de carboidrato mais comum produzida pela planta durante a fotossíntese e é armazenada no vacúolo das células dos tecidos não fotossintéticos. A sacarose é extraída dos colmos da planta, onde é mais abundante, para ser usada como alimento, conhecida como açúcar de mesa e como matéria prima para obter outros subprodutos a partir da fermentação e destilação tais como o álcool que é vendido no Brasil como “cachaça” uma bebida popular entre os brasileiros e como biocombustível, sendo uma alternativa energética amigável com o meio ambiente ao reduzir as emissões de CO<sub>2</sub>, quando comparado com o uso da gasolina. A cana-de-açúcar é considerada uma das culturas agrícolas mais importantes no planeta, por ser uma fonte de alimento e de energia renovável, isso devido à sua capacidade de armazenar grandes quantidades de sacarose nos entrenós dos seus colmos como parte da maturação da planta ao longo do tempo. O cultivo da cana-de-açúcar para extração da sacarose, resulta em um aumento na economia convertendo a indústria sucroalcooleira como um dos setores mais modernos e competitivos no mundo, no Brasil, a produção de sacarose tem tido importância destacada, sendo um pilar na economia do país, situando-lhe como líder mundial em produção durante vários anos consecutivos.

Devido à grande importância econômica da cana-de-açúcar pela extração da sacarose, este projeto visa ao entendimento molecular dos genes que dão lugar à expressão da enzima, sacarose 6-fosfato fosfohidrolase (S6PP), responsável pela liberação da sacarose no último passo da síntese. Pretende-se entender a função de dois possíveis genes homólogos *S6PP* identificados no genoma da cultivar híbrida de *Saccharum spp.*, variedade R570, onde verificou-se a existência de uma duplicação gênica em tandem (sequência denominada como *S6PP*) que resulta na predição de transcrição de dois peptídeos iguais, há evidências que essa duplicação é compartilhada com outros genomas de plantas (Phytozome,11). A relevância que introduzimos neste trabalho é a proposta de uma nova isoforma desse gene nunca publicada nem evidenciada em outros trabalhos, que sugere a conjugação destas duas sequências em tandem em que os genes conjugados são preditos a transcrever um único transcrito contendo os dois domínios da S6PP isso decorre possivelmente de *splicing* alternativo, (esta isoforma é denominada S6PP\_2D). A partir dessa descoberta surge a hipótese de que *Saccharum spp.* apresenta duas

variantes do gene *S6PP*, uma compartilhada com as demais espécies e outra resultante de um transcrito alternativo com dois domínios funcionais da *S6PP*, e o objetivo principal deste trabalho é elucidar a função de dois genes homólogos *S6PP* identificados no genoma da cultivar R570 de cana-de-açúcar. Para dar resposta à nossa hipótese introduzimos no trabalho de dissertação os conceitos principais da importância econômica da cana-de-açúcar, o metabolismo da sacarose e a descrição dos genes *S6PP* que são o foco principal. Realizamos estudos que compreendem análises moleculares para o entendimento do entorno genômico onde os genes *S6PP* estão inseridos, assim como do perfil da expressão desses genes durante o desenvolvimento da planta. A função bioquímica das proteínas *S6PP* foi estabelecida mediante expressão em sistemas heterólogos, purificação, e ensaios enzimáticos.

### 1.1 A sacarose

A sacarose (fórmula  $C_{12}H_{22}O_{11}$ ) é um dissacarídeo de 12 carbonos composto de unidades únicas de glicose e frutose (monossacarídeos), que são ligadas por uma ligação glicosídica (KITTS, 2010), é um tipo de açúcar que está presente naturalmente em todas as frutas e legumes. Os açúcares representam a principal fonte de energia para todos os organismos eucarióticos. As plantas são autotróficas e utilizam energia da luz solar para converter o dióxido de carbono ( $CO_2$ ) em carboidratos que são essenciais para processos fundamentais no crescimento da planta (SULPICE et al., 2009). A clorofila nas folhas das plantas capta a energia solar e permite a fixação do  $CO_2$  atmosférico absorvido pelos estômatos presentes nas folhas. A água absorvida pelas raízes no solo é dissociada em próton e oxigênio, produzindo o poder redutor e ATP necessários à fixação. Este processo é denominado fotossíntese (LORIMER; BUCHANAN; WOLOSUK, 2010;). A sacarose é a principal forma de carboidrato formado durante a fotossíntese e seu metabolismo fornece hexoses como glicose, frutose e outros açúcares que são necessárias para gerar energia e sintetizar celulose, amido, frutanos e compostos antioxidantes (KITTS, 2010; RUAN, 2014). Também funciona como uma reserva de carbono para as funções metabólicas e como metabolito de sinal em plantas modulando a expressão de genes que codificam proteínas transportadoras e de reserva, afetando

também a expressão de genes envolvidos na divisão celular e diferenciação, desempenha papéis cruciais no crescimento, desenvolvimento, armazenamento, transdução de sinal e aclimação às tensões ambientais nas plantas (LUNN; MACRAE, 2003; JIANG, et al., 2015).

Portanto, o transporte e partição de açúcares de folhas fototróficas (fonte) para órgãos heterotróficos (dreno) através do floema são os principais parâmetros que controlam a produtividade das culturas (AINSWORTH; BUSH, 2010). A exportação de sacarose das folhas de origem é rigidamente regulada pela atividade fotossintética, o metabolismo do amido nos cloroplastos e o armazenamento transitório de açúcares dentro dos vacúolos. Além disso, fatores ambientais, tanto bióticos quanto abióticos, também afetam a alocação de açúcares para a raiz e microrganismos da rizosfera (LEMOINE et al. 2013).

## 1.2 Metabolismo da sacarose

O metabolismo da sacarose é constituído de três grandes etapas: (i) **síntese** de sacarose como parte da fotossíntese, (ii) **transporte** de sacarose livre através dos vasos condutores do floema para ser difundido na planta inteira e (iii) **armazenamento** de sacarose nos vacúolos das células do parênquima dos órgãos dreno, incapazes de realizar a fotossíntese (Figura 1) (LORIMER; BUCHANAN; WOLOSUK, 2010; WANG, et al., 2013)

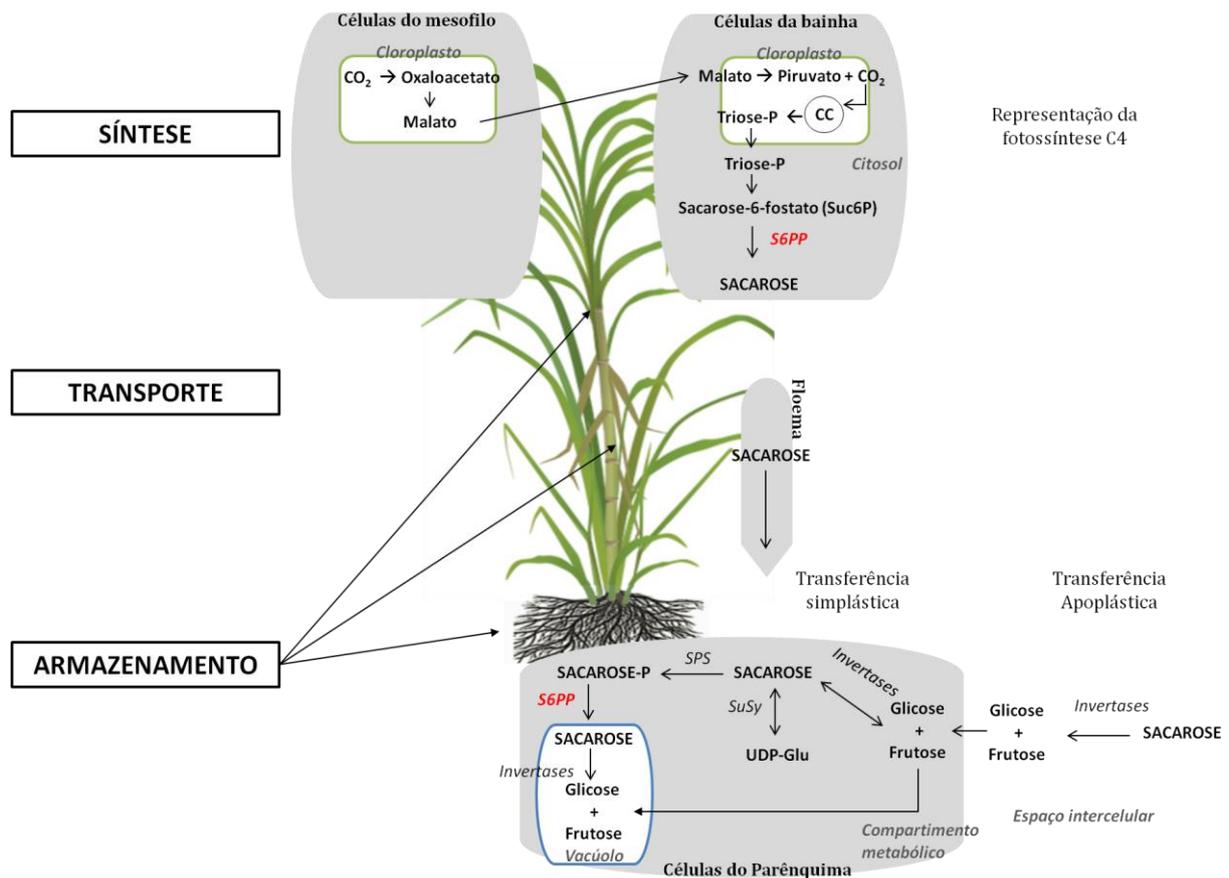


Figura 1. **Metabolismo geral da sacarose.** Consiste em três grandes etapas; Síntese da sacarose, transporte da sacarose livre através dos vasos condutores do floema e o armazenamento da sacarose nos vacúolos das células do parênquima dos órgãos dreno, incapazes de realizar a fotossíntese. Fonte: Modificado de WANG et al. (2013).

### 1.2.1 Fotossíntese em cana-de-açúcar; metabolismo C4

A atividade fotossintética na cana-de-açúcar, uma planta com metabolismo C4 é dividida entre dois tipos de células que tem cloroplastos, células do mesófilo e células da bainha vascular (anatomia *Kranz*), estas células encontram-se muito próximas umas das outras e conectadas por uma rede de plasmodesmas que fornece uma rota que facilita o fluxo de metabolitos entre os dois tipos de células. Na primeira etapa do processo ocorre a conversão de energia radiante em energia química para serem formadas duas moléculas de alta energia, chamadas NADPH (nicotinamida adenina dinucleótido fosfato) e ATP (adenosina trifosfato), estas são produzidas pela dissociação das moléculas de água em prótons ( $\text{H}^+$ ) e oxigênio ( $\text{O}_2$ ). Na segunda etapa, a planta utiliza a energia armazenada nesses compostos para

converter o dióxido de carbono ( $\text{CO}_2$ ) em moléculas de gliceraldeído 3-fosfato, também chamado de Triose-P, um intermediário para diversas vias metabólicas. A fixação do  $\text{CO}_2$  na fotossíntese  $\text{C}_4$  começa nas células do mesófilo e envolve a carboxilação do fosfoenol piruvato (PEP) com o  $\text{CO}_2$  pela enzima fosfoenol piruvato carboxilase (PEPC), formando oxaloacetato (OAA) um composto de quatro carbonos (daí o nome  $\text{C}_4$ ). O OAA é metabolizado a malato, este é transportado para dentro da bainha vascular onde é descarboxilado a piruvato gerando uma alta concentração de  $\text{CO}_2$ . O piruvato é transportado de novo à célula do mesófilo, onde o aceptor inicial de  $\text{CO}_2$ , o PEP, é regenerado pela enzima piruvato fosfato diquinase (PPDK). O  $\text{CO}_2$  é fixado e convertido em carboidrato através do Ciclo de Calvin (CC) que também ocorre nos cloroplastos das células da bainha e possui três fases: carboxilação, redução e regeneração. Na primeira fase o  $\text{CO}_2 + \text{H}_2\text{O}$  são incorporados à molécula ribulose 1,5-bifosfato (RuBP), esta reação de carboxilação é catalisada pela enzima ribulose bifosfato carboxilase (Rubisco), o composto de seis carbonos resultante é muito instável e rapidamente é dissociado em duas moléculas de três carbonos (3-fosfoglicerato). A próxima fase deste ciclo inclui a formação de gliceraldeído-3-fosfato (Triose-P) usando a energia obtida a partir do NADPH e do ATP gerados durante a fase de captação da energia luminosa. Já, na terceira fase algumas das Triose-P formadas são usadas dentro do cloroplasto para síntese de amido e são de novo incorporadas ao ciclo para regenerar o aceptor RuBP e assim continuar com o ciclo e outras das Triose-P são liberadas ao citosol para sintetizar a sacarose (CHRISPEELS; SADAVA, 2003; LORIMER; BUCHANAN; WOLOSUK, 2010).

A enzima Rubisco que forma parte do ciclo de Calvin é também capaz de catalisar uma reação simultânea e oposta de oxigenação da RuBP em um processo chamado fotorrespiração, esta resulta em perda de  $\text{CO}_2$  nas células que estejam fixando  $\text{CO}_2$  por meio do ciclo de Calvin. As plantas  $\text{C}_4$ , geram uma alta concentração de  $\text{CO}_2$  durante a descarboxilação do malato, o que suprime a oxigenação do RuBP, inibindo a fotorrespiração.

### 1.2.2 Síntese da sacarose em plantas

Este carboidrato assim como o amido são produzidos a partir da molécula de Triose-P gerada pelo ciclo de Calvin. Tanto as plantas C3 como as plantas C4 apresentam uma via metabólica da sacarose igual.

A síntese da sacarose ocorre no citosol e começa com a união de duas Triose-P que saem do cloroplasto mediante o transportador de Triose-P. Já, no citosol, a Triose-P é convertida em frutose 1,6-bifosfato pela enzima aldolase e depois é hidrolisado para frutose-6-fosfato (Frut6P) pela enzima frutose 1,6- bifosfatase. A Frut6P é então convertida para glucose 6-fosfato (Glc6P), esta última é convertida para glucose 1-fosfato e finalmente é formado UDP-glucose via UDP-glucose pirofosforilase. A síntese da sacarose é controlada pela atividade da sacarose fosfato sintase (SPS) (EC 2.4.1.14) que catalisa a reação [UDP-glucose + Frut6P → UDP + Sacarose 6-fosfato (Suc6P)] (Figura 2-A). O Suc6P é o substrato da enzima Sacarose 6-fosfato fosfohidrolase (S6PP ou SPP) (EC 3.1.3.24) que realiza uma reação de hidrólise que remove o Pi e deixa a sacarose livre no citosol [Suc6P + H<sub>2</sub>O → Suc + Pi] (Figura 2-B). Essa sacarose livre será translocada para os tecidos dreno e o Pi retorna ao cloroplasto mediante um transportador de fosfato para continuar com o abastecimento de Triose-P para o citosol (DINARDO-MIRANDA; DE VASCONCELLOS; DE ANDRADE LANDELL, et al. 2008; SMITH et al. 2010).

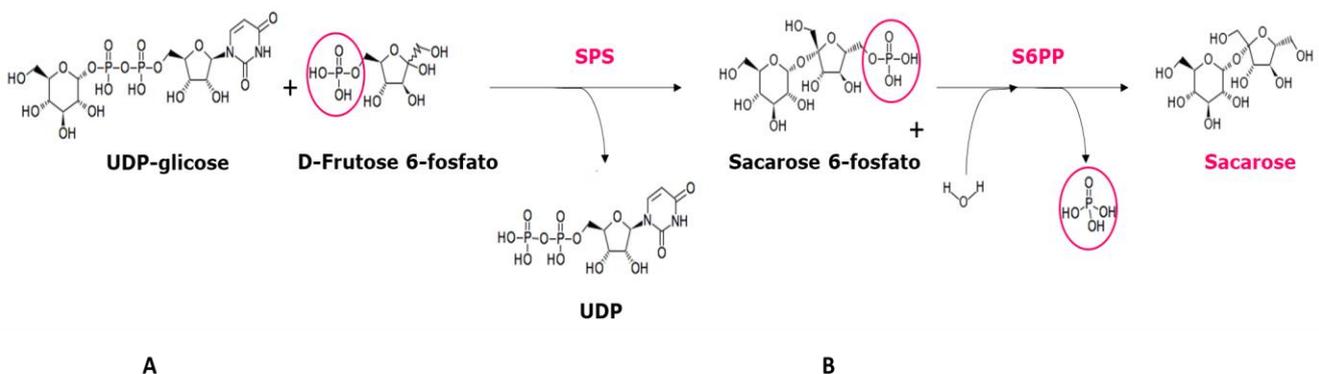


Figura 2. **Últimos passos da via metabólica da síntese de sacarose.** A) Reação enzimática da enzima Sacarose fosfato sintase, onde é obtida a sacarose 6-fostato e uma molécula de UDP, mediante a catálise do substrato UDP-glucose e frutose 6-fosfato. B) Reação enzimática da enzima Sacarose 6-fosfato fosfohidrolase, neste passo ocorre a liberação da sacarose, após uma reação de hidrólise onde o substrato sacarose 6-fosfato é desfosforilado para obter a molécula de sacarose e uma molécula de fosfato inorgânico.

### 1.2.3 Transporte da sacarose através do floema

O floema é um componente central do complexo sistema vascular das plantas terrestres e desempenha um papel vital na movimentação de fotoassimilados de locais de aquisição primária para os tecidos e órgãos heterotróficos da planta. De fato, até 50-80% do CO<sub>2</sub> fotoassimilado em uma folha madura é transportado para fora da folha no floema para suprir as necessidades dos órgãos não fotossintéticos da planta (KALT-TORRES et al., 1987; AINSWORTH; BUSH, 2010).

A exportação de carboidratos das folhas fotossintetizadoras (fontes) fornece o substrato para o crescimento e manutenção de tecidos vegetais não-fotossintéticos (dreno), e o floema é o sistema de entrega de açúcares exportados. Altas concentrações de açúcares, predominantemente sacarose, nos elementos crivados do floema elevam a pressão de turgescência, resultando em pressão hidrostática impulsionada pelo fluxo dos açúcares. Já, nos tecidos dreno, onde os açúcares são descarregados a pressão de turgescência cai. Este transporte de floema controlado pela sacarose é o caminho predominante através do qual todos os recursos, incluindo nutrientes, água e moléculas de sinalização, são importados para os órgãos dreno, incluindo os meristemas de raiz e apical. No entanto, existe considerável flexibilidade no carregamento do floema entre as espécies, e nem todas as espécies aderem estritamente a uma estratégia (RENNIE; TURGEON, 2009; SLEWINSKI; BRAUN, 2010; AINSWORTH; BUSH, 2010; RUAN *et al.*, 2012; RUAN, 2014).

### 1.2.4 Armazenamento de sacarose

A sacarose é armazenada nas células do parênquima dos tecidos não fotossintetizantes, podendo entrar por transferência simplástica como sacarose. Ou mediante transferência apoplástica, aqui as reações ocorrem nos espaços externos do parênquima e a sacarose é quebrada em glicose e frutose. Uma vez formadas essas hexoses, irão penetrar no compartimento metabólico do parênquima por um processo de simples difusão. No citosol essas hexoses sofrem um processo de interconversão para formar de novo uma molécula de sacarose, e essas moléculas de sacarose formadas e aquelas que entraram por via simplástica, serão convertidas em UDP-glicose pela ação regulatória da enzima Sacarose sintase (SuSy), EC 2.4.1.13, numa reação reversível de acordo as necessidades fisiológicas da planta.

Parte dessa sacarose é fosforilada pela enzima SPS para ser ativada formando o intermediário Suc6P que é desfosforilado pela enzima S6PP antes da entrada no vacúolo (HATCH, 1964; JIANG et al., 2015), a quebra da ligação fosfato fornece a energia necessária para a entrada no vacúolo, onde é acumulada (Figura 1) (DINARDO-MIRANDA et al. 2008).

Um exemplo interessante são os colmos da cana-de-açúcar. Estes destacam-se pela alta capacidade de acúmulo de sacarose. O colmo é organizado por uma sucessão de entrenós em diferentes estágios fisiológicos que acumulam a sacarose em condições favoráveis de maturação. Os colmos tendem a acumular a maior quantidade de sacarose perto da última fase do ciclo da cultura, quando tem baixa taxa de crescimento. A temperatura também é um dos fatores que interfere no armazenamento, pois, uma temperatura natural de maturação, abaixo de 21°C faz com que a sacarose seja desviada para ser armazenada nos colmos, e assim, restringir o crescimento vegetativo (ROBERTO, 2015).

### **1.3 Cana-de-açúcar**

Até a presente data, a maior concentração de sacarose pura encontrada em plantas é observada em cana-de-açúcar e beterraba sacarina. A sacarose foi extraída pela primeira vez da cana-de-açúcar (*Saccharum robustum*) há mais de 12.000 anos em Papua Nova Guiné (KITTS, 2010). A cana-de-açúcar (*Saccharum spp.*) é uma monocotiledônea de metabolismo C4, que apresenta grande acúmulo de sacarose no colmo, o que a torna uma importante cultura econômica. Pertence à família das gramíneas (Poaceae), ordem poales, subfamília Panicoidae, da tribo Andropogoneae que inclui os cereais dos gêneros *Sorghum* (sorgo) e *Zea mays* (milho), subtribo Saccharininea (GRIVET; ARRUDA, 2002; KUMAR; TRIPATHI, 2015). Acredita-se que é originária do Sudeste Asiático, na grande região das ilhas da Nova Guiné e Indonésia (DANIELS; ROACH, 1987). O gênero *Saccharum* forma um grupo de cruzamentos estreitamente relacionados e compreende seis espécies, duas espécies selvagens; *S. robustum* e *S. spontaneum* e quatro grupos de clones cultivados antigamente; *S. officinarum*, *S. edule*, *S. barberi* e *S. sinense* (D'HONT; GLASZMANN, et al. 1998; D'HONT et al. 2001).

O desenvolvimento de programas de melhoramento genético permite a expansão de culturas resistentes e melhor adaptadas às diversas condições ambientais, que são cultivadas para produção de sacarose. Variedades modernas de cana-de-açúcar são híbridos interespecíficos (*Saccharum spp.*) selecionados após intensivo cruzamento e seleção dentro do gênero *Saccharum*, envolvendo principalmente cruzamento entre as espécies *Saccharum officinarum* L. e *S. spontaneum* L. A espécie *S. officinarum* acumula níveis muito elevados de sacarose na célula, porém apresenta baixa resistência a doenças. Inversamente, *S. spontaneum* não acumula sacarose no colmo, porém é uma espécie com um elevado nível de resistência a doenças, adaptabilidade e tolerância a estresse (SREENIVASAN; AHLOOWALIA; HEINZ, et al. 1987; D'HONT, et al. 2001).

Os cultivares modernos híbridos resultantes, apresentam genomas extremamente complexos com elevado número de cromossomos, que pode variar entre  $2n=100-120$ , sendo 70-80% de *S. officinarum*, 10-23% de *S. spontaneum*, dependendo do cultivar (D'HONT, et al. 2001). A compreensão do genoma dessa cultura agrícola ainda é limitada. Possui um arranjo genético complexo, com várias cópias de cada cromossomo correspondendo a múltiplos alelos. Desta forma, torna-se difícil o entendimento sobre como características genéticas são transferidas e como funcionam as múltiplas variantes de cada gene na planta, dificultando o melhoramento e a obtenção de variedades mais produtivas de cana-de-açúcar. Alguns exemplos serão descritos a seguir.

### **1.3.1 Cultivares híbridos R570 e SP80-3280**

Cultivares modernos híbridos foram desenvolvidos mediante vários ciclos de cruzamentos e seleção entre as espécies *S. officinarum* e *S. spontaneum*, afim de obter variedades mais produtivas em sacarose e ao mesmo tempo resistentes a doenças. Dentre esses híbridos, as variedades R570 e SP80-3280 são utilizadas em programas de melhoramento e também comercializadas para cultivo e extração de sacarose no Brasil.

O cultivar R570 foi desenvolvido na ilha francesa de "Réunion", e selecionado por apresentar boas características agrônômicas tais como, fácil adaptação ao solo e boa germinação o que conduz a grandes extensões de cultivo e produção,

facilidade na colheita mecânica e por apresentar um alto conteúdo de fibras e açúcar (eRcane). Possui um genoma de 10.000 Mb e  $2n=107-115$  que compreende aproximadamente 80% de cromossomos de *S. officinarum* e 10% de *S. spontaneum* sendo 10% restante o resultado de recombinação cromossômica interespecífico (D'HONT et al., 1998; DE SETTA, et al., 2014). Esse cultivar foi usado na construção da biblioteca comercial de clones genômicos de BACs (do inglês, *Bacterial Artificial Chromosome*), de cana-de-açúcar (TOMKINS et al., 1999).

A variedade SP80-3280 foi desenvolvida no Brasil pelo Centro de Tecnologia Canavieira (CTC). Foi selecionada por se destacar no alto teor de sacarose e alta produtividade em soqueira, apresenta crescimento inicial vigoroso o que faz florescer, apresenta pouca isoporização o que facilita a extração da sacarose, o teor de fibra é alto, tem uma exigência média de solo fértil (SOCICANA, consultado em 2018) e é muito tolerante a herbicidas. A SP80-3280, das variedades SP, é mais indicada para programas de melhoramento e tem sido cultivada por pelo menos há 22 anos (FERREIRA et al., 2010; OKURA et al., 2012). Este cultivar foi utilizado nas bibliotecas de cDNA do projeto de sequenciamento de EST (do inglês *Expressed Sequence Tag*) de cana-de-açúcar, SUCEST (*Sugarcane EST*) (VETTORE et al., 2001).

Os dois cultivares híbridos mencionados acima, assim como as espécies usadas nos cruzamentos (*S. officinarum* e *S. spontaneum*) foram selecionados neste trabalho como organismos de estudo da expressão dos genes *S6PP*.

### **1.3.2 A cana-de-açúcar no mundo**

Em 2016 foram plantados 26.774.304 milhões de hectares de cana-de-açúcar em regiões tropicais e subtropicais em mais de 100 países produtores ao redor do mundo, destacando, Brasil, Índia, China, Tailândia, Paquistão, México, Colômbia, Austrália, Estados Unidos e Indonésia, produzindo no total aproximadamente 1.890.661.751 toneladas (figura 3) (FAOSTAT, 2018).

O número de plantações de cana-de-açúcar no mundo tem aumentado na última década, e em 2018 o Brasil continua liderando como o maior produtor mundial de cana-de-açúcar (FAOSTAT, 2018).

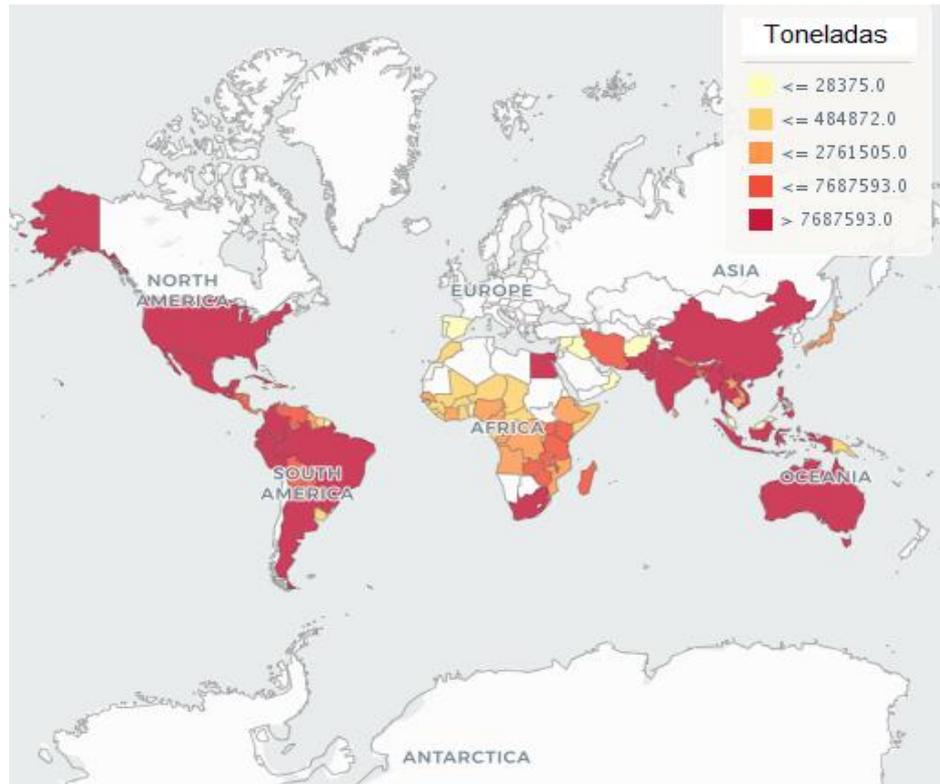


Figura 3. **Quantidades de produção em toneladas de cana-de-açúcar por país.** (FAOSTAT,2018)

### 1.3.3 Distribuição dos cultivos de cana-de-açúcar no Brasil

Dentro do território brasileiro a cultura ocupa diferentes regiões, principalmente no centro-sul e nordeste, adaptando-se muito bem às condições climáticas. No primeiro levantamento da safra 2018/2019 a área cultivada destinada à atividade sucroalcooleira foi de aproximadamente 8.661.4 mil hectares (ha), dos quais 4476.3 mil ha (51.7%) localizam-se no estado de São Paulo (SP) sendo o maior produtor nacional, seguido por Goiás (GO) com 917.1 mil ha (10.6%), Minas Gerais (MG) com 851.3 mil ha (9.8%), Mato Grosso do Sul (MS) com 664.7 mil ha (7.7%), Paraná (PR) com 570.3 mil ha (6.6%), Alagoas (AL) com 286.2 mil ha (3,3%), Mato Grosso (MT) com 235.3 mil ha (2,7%), Pernambuco (PE) com 231 mil ha (2,6%) e Paraíba (PB) com 119.7 mil ha (1.4%) totalizando 96% da área total cultivada, o restante é dividido entre outros estados pouco produtores, nos que destacam, Rio Grande do Norte (RN), Espírito Santo (ES), Sergipe (SE), etc. totalizando a produção de cana no Brasil (CONAB, 2018).

### 1.3.4 Importância econômica

Devido à extraordinária capacidade da cana-de-açúcar de armazenar sacarose nos entrenós dos seus colmos, ela tornou-se uma importante cultura na agroindústria. Segundo MARAFANTE (1993), a cana-de-açúcar contém cerca de 90% de suco, do qual se pode extrair entre 10 e 20% de açúcar, o que a torna a planta que produz mais açúcar no mundo. A partir desse cultivo é possível produzir 80% do açúcar que é distribuído no globo (DE SETTA et al., 2014). A importância econômica da cana-de-açúcar está vinculada a três agroindústrias: açúcar, álcool e aguardente, além disso, como produto secundário há também a produção de bioeletricidade que alimenta a região na proximidade das usinas (OMETTO, 2000). A indústria sucroalcooleira é um dos setores mais modernos e competitivos no mundo.

Desde 1532, ano em que a cana-de-açúcar foi introduzida no Brasil, ela tem tido importância destacada, sendo um pilar na economia do país. Essa produção destina-se às fabricas de etanol (anidro e hidratado) e às de açúcar (refinado, cristal e demerara). Desde então, avanços importantes nos processos de cultivo e industrialização da cana foram desenvolvidos, levando à cultura para um elevado nível tecnológico. Cerca de 50 anos atrás, além de açúcar, as usinas passaram a produzir etanol, este biocombustível tem um papel ambiental muito importante, como alternativa energética para reduzir a emissão de gases causadores do efeito estufa, reduzindo em 70% a emissão de CO<sub>2</sub> na atmosfera em relação à gasolina (UNICA, 2015).

Na safra 2018/2019 a produção de açúcar no Brasil consolidou-se em 34.245,9 mil toneladas e o etanol chegou a ser produzido em 30,4 bilhões de litros. A maior produção de etanol total continua concentrada na região Sudeste, com 57.7% do total produzido no país, seguido pelo Centro-Oeste (31%), Nordeste (5.7%), Sul (4.9%) e Norte (0.8%) (CONAB, 2018). No primeiro semestre de 2016 foram exportadas aproximadamente 12.504.455 toneladas de açúcar (que corresponde aproximadamente a 50% de toda a exportação mundial de cana) do Brasil para mais de 100 países que gerou aproximadamente US\$ 3.928 milhões em receita. As exportações de álcool figuraram em receita com US\$ 520,090 milhões e em geral o complexo sucroalcooleiro gerou em receita US\$ 4.455 milhões (AGROSTAT, 2016). Devido a sua importância econômica nacional no setor sucroalcooleiro, são

necessários investimentos em pesquisa para atender a demanda global para os próximos anos e assim continuar liderando o mercado global nas exportações de produtos e subprodutos obtidos a partir da cana de açúcar.

Devido ao alto valor econômico da cana-de-açúcar como “commodity” mundial na produção de açúcar (sacarose), energia renovável e outras biomoléculas, este trabalho visa o entendimento molecular dos genes que levam à expressão da enzima sacarose 6-fosfato fosfohidrolase (S6PP), responsável pela liberação de sacarose.

#### 1.4 Enzima sacarose 6-fosfato-fosfohidrolase (S6PP)

A enzima sacarose 6-fosfato fosfohidrolase (S6PP), EC 3.1.3.24, PF 08472, catalisa a reação do último passo da via metabólica da síntese de sacarose, na qual, a sacarose-6-fosfato (Suc6P), produto da reação enzimática da sacarose fosfato sintase (SPS; EC 2.4.1.14), é desfosforilada e liberada como sacarose. No entanto, também poderia participar no preparo da sacarose prévio à entrada no vacúolo onde será acumulada. O Suc6P é altamente específico para a S6PP, a reação é essencialmente irreversível e direciona o fluxo da reação catalisada pela SPS para síntese final de sacarose livre (Figura 4) (LUNN et al., 2000; FIEULAINÉ; LUNN; FERRER et al., 2007)

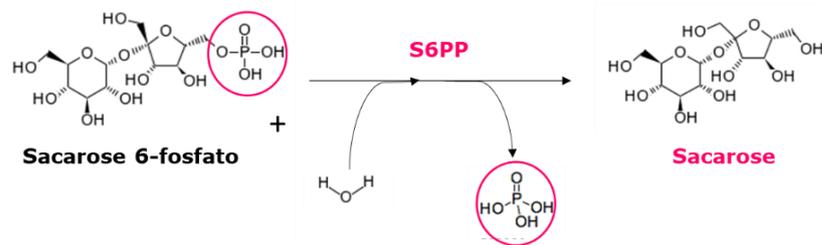


Figura 4. **Reação catalisada pela sacarose 6-fosfato fosfohidrolase (S6PP).** O substrato sucrose-6-fosfato é hidrolisado no carbono 6 da molécula de frutose, deixando sacarose livre e liberando um fosfato inorgânico.

Esta enzima (S6PP) é encontrada em espécies representativas dos principais grupos de plantas vasculares e não vasculares: algas verdes, plantas hepáticas, musgos, gimnospermas e angiospermas, identificada a partir de tecidos foliares

fotossintetizantes. A sacarose é também sintetizada por algumas cianobactérias marinhas e de água doce, como resposta de adaptação para manter o balanço osmótico causado pelo estresse derivado do acúmulo de sal ou estresse derivado das diferenças de temperatura e assim estabilizar a sua estrutura para desenvolver as funções vitais da bactéria (LUNN, 2002; BUT et al., 2013). Segundo o levantamento bibliográfico sobre S6PP conforme resumido e listado na tabela 1, em plantas das que se tem registro de estudos, esta enzima é descrita como homodímero com peso molecular em torno de 120 KDa. A enzima S6PP da folha do arroz é um homodímero de duas subunidades de 50 KDa, a S6PP de folha de milho tem uma massa molecular entre 47 e 49 KDa e a S6PP da cianobactéria *Synechocystis* sp. tem entre 27-28 KDa, sendo menor do que a S6PP das plantas superiores, mas, apresentam significativa similaridade na sequência de aminoácidos (LUNN et al., 2000; LUNN, 2002).

Tabela 1. **Resumo de estudos da enzima S6PP realizados em plantas e bactérias.**

Organismo	Tamanho da enzima	Fonte da extração	Estado da purificação	Estrutura	Referencias
Ervilha ( <i>Pisum sativum</i> )	120 kDa ( $\pm 10$ kDa) (dímero 55 kDa) *	Semente	pura	Não determinada	WHITAKER, 1984
Arroz ( <i>Oryza sativa</i> L.)	120 kDa	Folha	Parcialmente pura	Não determinada	ECHEVERRIA; SALERNO, 1994
Arroz ( <i>Oryza sativa</i> L.)	100 kDa (dímero 50 kDa) *	Folha	pura	Não determinada	LUNN et al., 2000
Milho ( <i>Zea mays</i> ) SPP1	47,2 kDa (predição por aminoácidos e SDS-PAGE)	Expressão heteróloga	extrato	Não determinada	LUNN et al., 2000
Cianobacteria <i>Anabaena</i> sp.	26-28 kDa	Nativa/Expressão heteróloga	pura	Determinada	CUMINO; EKEROTH; SALERNO, et al., 2001
Cianobacteria <i>Synechocystis</i> sp.	27 kDa	Expressão heteróloga	pura	Não determinada	LUNN, 2002
Cianobacteria <i>Synechocystis</i> sp.	28 kDa	Expressão heteróloga	pura	Determinada	FIEULAINÉ et al., 2005
Cianobacteria <i>Synechocystis</i> sp.	28 kDa	Expressão heteróloga	pura	Determinada	FIEULAINÉ et al 2007
Proteobacteria <i>Methylobacillus flagellatus</i>	336 KDa (tetrâmero 84 kDa) *	Expressão heteróloga	pura	Não determinada	BUT et al., 2013
<i>Arabidopsis thaliana</i>	$\approx 90$ kDa (dímero 52 kDa) *	Expressão heteróloga	pura	Não determinada	ALBI et al., 2016

\*Cada subunidade proteica corresponde ao tamanho indicado dentro do parêntese.

### 1.5 Estudos prévios dos genes *S6PP* no Genomics and Transposable Elements Laboratory (GaTE Lab)

Os genes *S6PP*, que codificam para a enzima S6PP, foram identificados no genoma da cultivar R570 de *Saccharum spp.* que foi sequenciado pelo laboratório, GaTE, a partir da biblioteca de fragmentos genômicos clonados em vetores BACs gerada por TOMKINS et al., (1999). O projeto do sequenciamento concluiu com a anotação e o depósito de 317 BACs por DE SETTA et al., (2014). Durante o processamento dos dados gerados do sequenciamento da R570, o grupo do GaTE Lab, liderado pela Professora Dra. Marie-Anne Van Sluys, verificou a existência destes genes (*S6PP*) em quatro destes fragmentos genômicos. Foi encontrada uma duplicação genica em tandem destes genes nas BACs denominadas SCHRba\_237\_G04 (Figura 5) e SCHRba\_096\_D24 correspondendo a possibilidade de transcrição de dois peptídeos com o mesmo domínio enzimático, mas, foi interessante notar que para outros dois clones de BACs denominadas SCHRba\_104\_G22 e SCHRba\_085\_J02, existe a predição de uma nova forma potencialmente duplicada nunca antes descrita que sugere a conjugação destas duas sequências em tandem em que os genes conjugados são preditos a transcrever um único transcrito, evento provavelmente decorrente de *splicing* alternativo, onde foi exonizado parte do espaço intergênico (Figura 5) de modo que ocorra uma proteína com dois domínios funcionais da S6PP. Esta duplicação é resultado da predição do software de anotação de genes chamado Caramujo®, que foi desenvolvido pelo Dr. Alessandro de Mello Varani do Departamento de Tecnologia da Faculdade de Ciências Agrárias e Veterinárias-UNESP em Jaboticabal, SP.

Assim, a duplicação em tandem encontrada pelo grupo não tinha sido publicada anteriormente em cana-de-açúcar, mas, é evidenciada em sequências genômicas do banco de dados de Phytozome 11, nas espécies *Sorghum bicolor* v3.1, *Setaria itálica* v2.2, *Panicum virgatum* v1.1 (DE SETTA et al., 2014; Phytozome 11) e recentemente durante este estudo foi encontrada a duplicação nos genomas de *Panicum hallii* v2.0, *Setaria viridis* v2.1 e *Aquilegia coerulea* v3.1 (Phytozome 11). Os genomas antes mencionados apresentam a duplicação do gene *S6PP*, com predição de transcrição de dois polipeptídios com a mesma função, assemelhando aos descritos acima (SCHRba\_096\_D24 e SCHRba\_237\_G04) da cultivar R570. Já,

no caso da duplicação do *S6PP* nos fragmentos genômicos SCHRBa\_104\_G22 e SCHRBa\_085\_J02) os genes duplicados são preditos a transcrever um único transcrito.

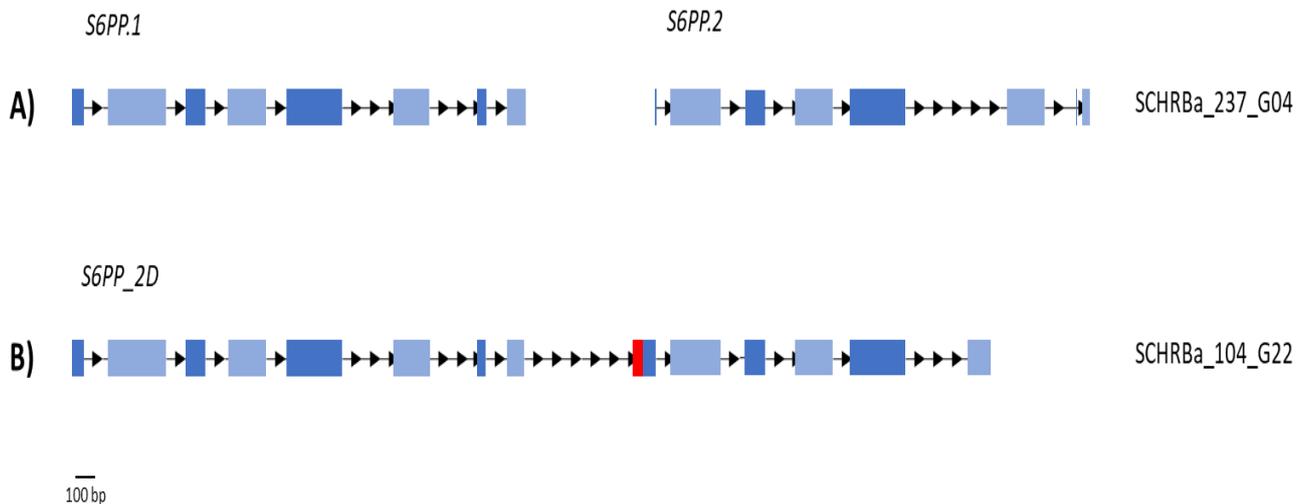


Figura 5. **Representação gráfica das isoformas dos genes *S6PP*, encontradas em dois diferentes fragmentos genômicos.** A) Duplicação genica em tandem, onde os genes são preditos a transcrever dois peptídeos com o mesmo domínio enzimático. B) Situação em que os genes *S6PP* são conjugados e são preditos a transcrever um único peptídeo que contém os dois domínios enzimáticos, parte do espaço intergenico foi exonizado (na cor vermelha). As setas pretas representam os introns e o sentido da transcrição, os quadros azuis representam os exons, o quadro vermelho representa o espaço intergenico que foi exonizado

## 2 HIPOTESE

H<sub>0</sub>: *Saccharum* spp. apresenta variantes do transcrito *S6PP*, uma sendo uma duplicação genica em tandem compartilhada com outras espécies e outra resultante de um transcrito alternativo que sugere a conjugação das sequências em tandem que contém dois domínios funcionais da S6PP para hidrólise do substrato sacarose-6-fosfato.

### 3 OBJETIVO

Elucidar a função de dois genes homólogos *S6PP* identificados no genoma da cultivar R570 de cana-de-açúcar.

Os objetivos específicos são listados a seguir:

1. Comparar a estrutura genômica dos genes *S6PP* dentro de fragmentos genômicos clonados em vetores BACs das variedades de cana de açúcar R570 e SP80-3280.
2. Determinar o perfil da expressão dos genes *S6PP* em cana-de-açúcar.
3. Estabelecer a função bioquímica das proteínas *S6PP* mediante a expressão em sistemas heterólogos e ensaios enzimáticos.

## 4 METODOLOGIA

### 4.1 ESTUDO MOLECULAR DO ENTORNO GENÔMICO DOS GENES *S6PP*

No intuito de entender o perfil da expressão dos genes *S6PP* na cana-de-açúcar em diferentes cultivares e diversos tecidos ao longo do tempo, foram realizados estudos de genômica comparativa, análise de RNAseq e quantificação de expressão.

#### 4.1.1 Genômica comparativa

Um estudo comparativo da estrutura genômica onde os genes *S6PP* estão inseridos foi desenvolvido. Fragmentos genômicos das variedades híbridas comerciais de cana-de-açúcar R570 e SP803280 (tabela 2) previamente clonados em vectores BACs e sequenciados, foram estudados para entender o ambiente genômico e o perfil de inserção de elementos de transposição (TEs, do inglês *Transposable Elements*) onde os genes *S6PP* são expressados. Pretendia-se responder as perguntas: o gene *S6PP* encontra-se em um único *locus*? A duplicação previamente identificada pelo grupo em R570 é compartilhada entre as duas variedades em estudo?

Todos os genes preditos durante a anotação das BACs foram analisados individualmente usando o banco de dados do *National Center for Biotechnology Information* (NCBI), já, os TEs foram analisados no banco de dados GIRM RePBASE®, considerando a regra 80-80-80 (WICKER et al., 2007). A partir disso, foi desenhada uma representação gráfica de cada uma das BACs que mostra as posições e o sentido da transcrição de cada gene predito. A genômica comparativa se deu com o alinhamento dos genes *S6PP* em tandem e paralelamente foi desenvolvida uma árvore filogenética tipo Maximum Likelihood modelo *T92+G*, 1000 replicas bootstrap, mediante o programa MEGA7, para realizar a árvore mencionada foi selecionada a região genômica em tandem contemplando o códon de início da primeira sequência e foi até o códon de terminação da segunda sequência duplicada incluindo exons, íntrons e região intergênica. O estudo ajudou para estudar a filogenia, possíveis agrupamentos e segregações entre os fragmentos.

Tabela 2. Lista dos fragmentos genômicos que contém os genes S6PP.

Cultivar de cana-de-açúcar	Nome da BAC	Tamanho da BAC (pb)
R570	085_J04	123.896
	096_D24	109.397
	104_G22	99.804
	237_G04	94.577
SP803280	109_H09	89.338
	149_N10	140.447
	228_N18	141.113
	273_J18	130.293
	257_A23	123.970
	258_B07	133,749
	264_H19	118.514
	264_P08	132.964
	441_A04	143.827
	453_B01	134.780
	456_J23	126.018
	465_H15	155.144
	480_A02	135.325
	492_H21	99.828
	494_M23	135.410
	524_N01	20.541
548_I17	126.419	

#### 4.1.2 Análise de expressão em bancos de dados públicos de RNAseq

Foi realizada uma análise da expressão dos CDSs (Do inglês, *Coding DNA Sequences*) preditos a partir dos genes *S6PP.1* e *S6PP\_2D* em três bancos de dados de RNAseq gerados a partir de bibliotecas de tecidos de cana-de-açúcar, as condições de desenvolvimento se mencionam a continuação:

- Banco de dados do laboratório GaTE, gerado de sequenciamento *paired end* HiSeq 1500 (Illumina™), a partir de amostras de meristema apical e raiz de três meses de idade da cultivar R570 de cana-de-açúcar (dados não publicados).

- Banco de dados gerado do sequenciamento *paired end* HiScanSQ (Illumina™), a partir de amostras da cultivar RB92-5345 infetadas e não infetadas com *Sporisorium scitamineum* (doença do carvão). Foram coletados dois estágios de infecção, os brotos de 5 dias depois após a infecção (5 DAI, pelas siglas em inglês) e a base do chicote, até 2 cm abaixo do colmo com 200 após a infecção (200 DAI), quando a doença é evidente (SCHAKER et al., 2016)
- Banco de dados gerado a partir do sequenciamento *single end* Illumina GAII, de amostras de plantas inteiras da variedade SP70-1143 infetadas com *Acidovorax avenae* subsp. *avenae* depois do enraizamento que foram coletadas depois de 7 dias de infecção (SANTA BRIGIDA et al., 2016).

O mapeamento dos *reads* nas sequências codificantes dos genes alvo foi realizado com o software *Bowtie2* (LANGMEAD, et al. 2009; LANGMEAD, 2010) mediante a plataforma Galaxy 2.2.6.2 (AFGAN, et al., 2016) disponibilizada no portal do Centro de Facilidade de Apoio a Pesquisa (CEFAP) da Universidade de São Paulo. Os parâmetros computacionais de alinhamento foram os já estipulados por *default* (padrão), sendo os mesmos para todos os mapeamentos. Estes parâmetros padrão envolvem o alinhamento sensível de todos os caracteres dentro de um *read* (*end-to-end*) e um determinado *read* poderia mapear apenas uma região da sequência de referência.

Já, a visualização dos alinhamentos foi mediante o *software Integrative Genomics Viewer 2.3* (IGV 2.3) (Broad Institute®, 2013), seguindo as especificações do fabricante, que mostra os alinhamentos dos *reads* mapeados contra o genoma de referência e a cobertura de *reads* na sequência alvo.

No intuito de quantificar e comparar a cobertura no mapeamento obtido, foram determinados os valores das expressões de normalização FPKM (do inglês *Fragments Per Kilobase Million*) usado para sequenciamento *paired-end* e o RPKM (do inglês, *Reads Per Kilobase Million*) usado para sequenciamento *single-end*.

### 4.1.3 Perfil da expressão dos genes *S6PP* em tecidos de cana-de-açúcar.

#### 4.1.3.1 Material vegetal e condições de cultivo

Os organismos usados neste experimento foram as variedades híbridas de cana-de-açúcar, R570 e SP80-3280 e as espécies *Saccharum officinarum* (variedade Badilla) e *S. spontaneum* (variedade IN84-58). Entrenós destas plantas maduras da coleção do GaTE foram colocados em substrato vegetal e vermiculita 1:0,75 complementado com fosfato e adubo para dar início à brotação de gemas. A germinação e o crescimento de novas plantas se deram em casa de vegetação do laboratório GaTE com luz natural, em condições ambientais do período correspondente a setembro de 2015 até a junho de 2016.

A coleta de material foi sempre realizada no final da tarde com presença de luz, a cada três meses de crescimento a partir da plantação, contemplando às idades de 3, 6 e 9 meses. As plântulas foram coletadas, lavadas com água corrente, separadas por tecidos: folha, raiz, meristema apical e colmo, foram congelados instantaneamente em nitrogênio (N<sub>2</sub>) líquido e guardadas a -80°C até serem processadas.

#### 4.1.3.2 Extração e tratamento do RNA total

Os tecidos (folha, raiz, meristema apical e colmo), das plantas de cana-de-açúcar, R570, SP80-3280, Badilla e IN84-58 com 3, 6 e 9 meses de idade, previamente coletados e armazenados em condições de congelamento foram macerados com *TissueLyser* II (QIAGEN®), com ciclos de frequência de oscilação de 30, durante 30 segundos, até a obtenção de um pó fino. Todo o processo foi sempre mantido em N<sub>2</sub> líquido.

Cada extração de RNA foi feita a partir de aproximadamente 100 mg de material vegetal macerado, adicionando 1 ml do reagente TRIzol® (Invitrogen®), a mistura foi suspensa rapidamente por agitação e congelada com N<sub>2</sub> líquido para ajudar na lise celular, foi descongelada com ajuda das mãos e incubada durante 5 minutos a temperatura ambiente. 200 µl de clorofórmio foram adicionados e misturados com vortex durante 15 segundos, se deixou repousando durante 10 minutos a temperatura ambiente. A mistura foi centrifugada a 12.000 g por 10

minutos a 4°C. As fases foram separadas pipetando cuidadosamente o sobrenadante e transferindo para um tubo novo, (sem pegar a interfase formada), 500 µl de isopropanol 100% foram adicionados no sobrenadante separado e foi incubado a -20°C durante a noite inteira. No dia seguinte os ácidos nucleicos foram sedimentados por centrifugação a 12.000 g por 10 minutos a 4°C, o sobrenadante foi descartado e o pellet formado foi lavado duas vezes com etanol 70% preparado com H<sub>2</sub>O tratada com dicarbonato de dietila (DEPC, pelas siglas em inglês) 0,01%, o etanol foi descartado com ajuda de ponteira para tirar a maior quantidade possível. Os *pellets* ficaram secando a 40°C durante 30 minutos para depois serem diluídos em 100 µl de H<sub>2</sub>O-DEPC. Foi feito um tratamento com DNase para remoção de DNA com o kit TURBO DNA-free (AM1907) (ambion®, Life Technologies™), seguindo as indicações do fabricante.

O RNA total livre de DNA foi quantificado por espectrofotometria com NanoDrop 1000 (UNISCIENCE®), seguindo as especificações do fabricante. A integridade do RNA total extraído e livre de DNA foi verificada em gel de agarose 2% em tampão TAE 1X-DEPC.

#### 4.1.3.3 *Transcrição reversa para síntese de cDNA*

Aproximadamente 3 µg de RNA total tratado com DNase foi utilizado para síntese da primeira fita de cDNA com transcriptase reversa usando o kit da enzima *SuperScript™ III Reverse Transcriptase* da Invitrogen®, seguindo as instruções do fabricante, num volume final de 20 µl. Os cDNAs sintetizados foram tratados com RNase H como sugerido no manual e foram armazenados a -20°C até serem usados.

#### 4.1.3.4 *Análise da qualidade da reação de síntese de cDNA*

A eficiência da síntese de cDNA foi verificada mediante uma reação de PCR qualitativa usando iniciadores que amplificam uma região de 100 bp do transcrito do gene Gliceraldeído-3-fosfato desidrogenase (GAPDH, pelas siglas em inglês) (ISKANDAR et al., 2004) sequencias apresentadas na tabela 3. Estes iniciadores servem como controle da eficiência da síntese cDNA.

Para as reações de PCR foi utilizado 1 µl de produto de cDNA com o kit da enzima *GoTaq<sup>®</sup> DNA Polymerase* (Promega<sup>®</sup>), 5 µl de Green Buffer 5X, 25 mM de MgCl<sub>2</sub>, 10 mM de dNTPs, 10 µM de *primers* e 1 u de enzima, numa reação de 25 µl. As condições de amplificação foram; um ciclo de desnaturação inicial a 95°C por 2 minutos, seguido de 35 ciclos de desnaturação a 95°C por 30 segundos, alinhamento dos *primers* a 60°C por 30 segundos e extensão a 72 °C por 1 minuto, para concluir com um ciclo de extensão final a 72°C por 7 minutos. Os produtos de PCR foram visualizados em gel de agarose 2% em tampão TAE 1X e a partir daí foram selecionadas as melhores amostras para serem usadas nos análises futuros.

Tabela 3. Lista de primers utilizados nos experimentos de PCR e qPCR.

SAS	Nome do primer	Sequência (5'→3')	Gene que amplifica
<b>Genes constitutivos</b>			
CA148161	Act_F	CCAGTTCATTGTCCACAAACAAG	Actina
	Act_R	TCTCCGGAATCCGTAGCAAA	
CA254672	GAPDH_F	CACGGCCACTGGAAGCA	GAPDH
	GAPDH_R	TCCTCAGGGTTCCTGATGCC	
SCCCST2001G02.g	PUB_F	CCCTCTGGTGTACCTCCAATTTG	Poliubiquitina
	PUB_R	CCGGTCCTTTAAACCAACTCAGT	
<b>Genes não constitutivos</b>			
	S6PP.1-2D_F	CTCAGCCAGAGAGGAAGCAG	S6PP.1 / S6PP_2D
	S6PP.1-2D_R	CACGTTTCTCCAACCTTCTGTG	
	S6PP-2D_F	ACACGTTTCATCTTGGAACCC	S6PP_2D
	S6PP-2D_R	ATCATAAGACGGGCTGAAGC	
	S6PP.1s-2Ds-BamHI_F	GGATCCATGGATAGGCTCGACGGTTCTG	S6PP.1 / S6PP_2D
	S6PP.1s-Sall_R	GTCGACCTACAAGATGAACGTGTGCTCGC	S6PP.1
	S6PP.2Ds-Sall_R	GTCGACTCAGCAGTCTGTTCTGTTGGGC	S6PP_2D

#### 4.1.3.5 Busca do melhor gene normalizador para quantificação da expressão por qPCR

Foi realizada uma busca do melhor gene constitutivo para ser usado como normalizador na técnica de qPCR, para isso, foram usados iniciadores que amplificam uma região de entre 90 e 100 pb dos transcritos dos genes Actina, GAPDH e Poliubiquitina (tabela 3), para este teste foram utilizadas as amostras de 3, 6 e 9 meses de idade de colmo e meristema apical, assim como raízes e folhas de 3 meses provenientes da variedade híbrida de cana-de-açúcar R570. Para a análise

de qPCR foi usado o kit SsoAdvanced™ Universal SYBR® Green Supermix, *BIO-RAD*, seguindo as instruções do fabricante.

#### 4.1.3.6 Desenho dos iniciadores para quantificação dos genes pela técnica de qPCR

Para quantificar a expressão do gene *S6PP* nas versões *S6PP.1* e *S6PP\_2D* pela técnica de qPCR, iniciadores específicos foram desenhados sobre as sequências de CDS preditas, o desenho envolve uma junção exon-exon para reduzir o risco de falsos positivos de qualquer DNA genômico contaminante. Estes iniciadores amplificam ampliações de 100 pb em regiões estratégicas dos CDS, os pares de oligonucleótidos específicos utilizados se encontram listados na tabela 3. O par de iniciadores *S6PP.1-2D* amplificam as duas sequências em estudo (*S6PP.1* e *S6PP\_2D*), já o par de primer *S6PP-2D* amplifica apenas a isoforma *S6PP\_2D* na região que foi exonizada (figura 6)

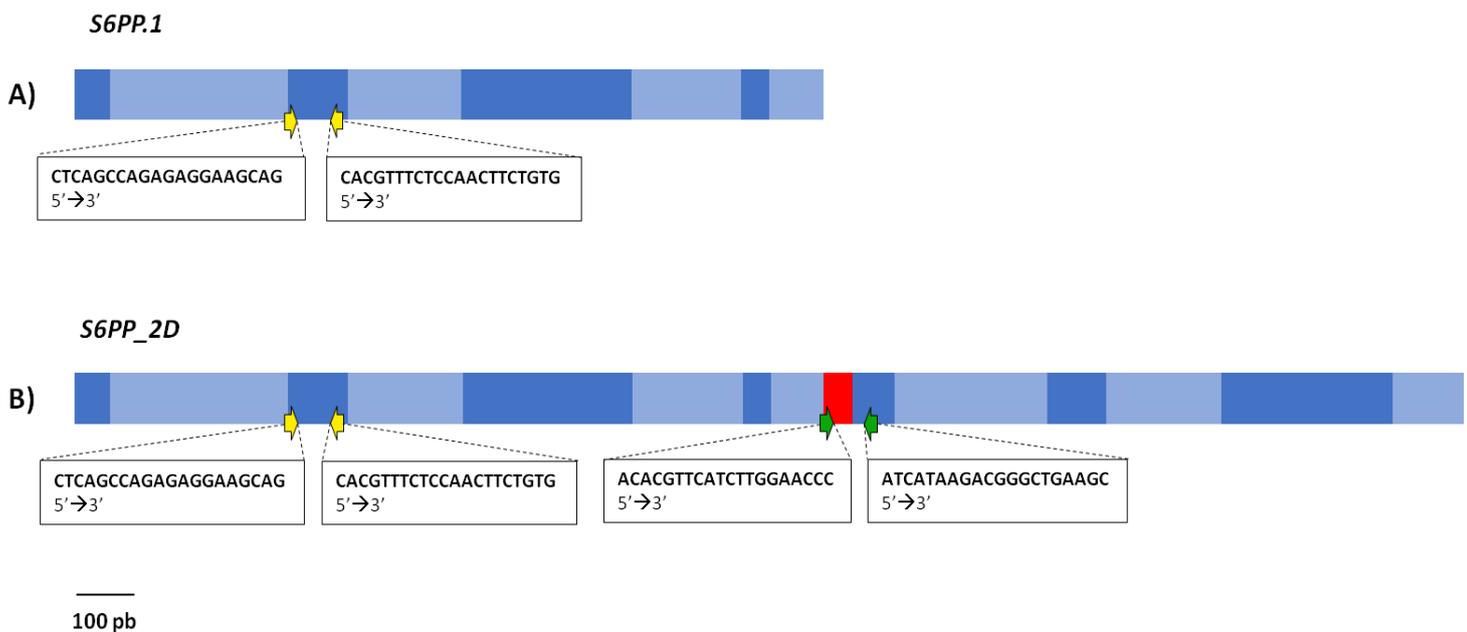


Figura 6. **Desenho de primers sobre os CDSs preditos dos genes *S6PP.1* e *S6PP\_2D* envolvendo a junção exon-exon.** A) Representação da sequência codificante do gene *S6PP.1*, B) Representação da sequência codificante do gene *S6PP\_2D*. Os retângulos azuis representam os exons. As setas amarelas representam o par de primers *S6PP.1-2D* que se alinha nas duas sequências, as setas verdes representam o par de primers *S6PP\_2D* que se alinha apenas na sequência *S6PP\_2D*.

#### 4.1.3.7 Análise quantitativo da expressão dos genes *S6PP* por qPCR

A reação de qPCR se deu usando o kit SsoAdvanced™ Universal SYBR® Green Supermix, *BIO-RAD*, seguindo as instruções do fabricante. Uma região dos transcritos dos genes *S6PP.1* e *S6PP\_2D* foram amplificadas usando os primers *S6PP.1-2D* e *S6PP.1* como explicado no item 4.1.3.6 e o gene constitutivo Poliubiquitina foi usado como normalizador da expressão genica. Já, os níveis de expressão foram analisados por quantificação comparativa com a formula  $2^{-dCt}$ . As amostras biológicas utilizadas em este experimento estão listadas na tabela 4.

Tabela 4. **Amostras utilizadas para análise da expressão por qPCR**

Cultivar	Idade	Tecido
R570 SP80-3280	3 meses	Folha
		Meristema
		Colmo
IN84-58 Badilla	6 meses	Raiz
		Meristema
	9 meses	Colmo
		Meristema
		Colmo

## 4.2 ESTUDO BIOQUÍMICO E FUNCIONAL DAS PROTEÍNAS *S6PP* EXPRESSADAS EM SISTEMAS HETERÓLOGOS

Com este estudo pretendia-se entender se as duas variantes da proteína *S6PP* seriam funcionais assim como conhecer a estrutura das proteínas mediante cristalografia em colaboração com o Grupo de Cristalografia, do Instituto de Física de São Carlos/USP, sob a orientação do professor Dr. Richard Charles Garrat.

### 4.2.1 Síntese das proteínas *S6PP* em linhagens de expressão de *E. coli*

#### 4.2.1.1 Preparo de DNA plasmidial em pequena escala (*MiniPrep*)

As sequencias dos genes *S6PP* nas duas versões *S6PP.1* e *S6PP\_2D* foram sintetizados pela empresa BlueHeron®, incorporando os sítios de restrição para as enzimas *XhoI* e *BamHI* no início e no fim da sequência respectivamente. Os genes

sintéticos denominados S6PP.1s e S6PP-2Ds foram recebidos dentro do plasmídeo pUCampMinusMCS em bactérias transformadas. A partir daí se procedeu à manipulação para a expressão heteróloga da enzima S6PP nas duas versões. O isolamento dos plasmídeos contendo os genes sintéticos deu início com a produção de clones, mediante o cultivo das bactérias em meio sólido Luria-Bertani (LB), contendo ampicilina como antibiótico de seleção do plasmídeo. No dia seguinte uma colônia isolada foi inoculada em 2 ml de meio líquido LB + ampicilina [ $100 \mu\text{g ml}^{-1}$ ] e foi incubada a  $37^\circ\text{C}$  a 230 rpm em shaker, Innova<sup>®</sup>44 (New Brunswick scientific) durante 18 horas. No dia seguinte 1 ml da cultura foi centrifugada  $14.000 \text{ g}$  durante 90 segundos na microcentrífuga MiniSpin<sup>®</sup> plus (Eppendorf<sup>®</sup>), o sobrenadante foi descartado e o precipitado celular foi suspenso em  $100 \mu\text{l}$  de solução I ( $50 \text{ mM}$  Tris-HCl,  $10 \text{ mM}$  EDTA,  $100 \mu\text{g ml}^{-1}$  RNase, pH 8.0) com ajuda de vortex. As células foram lisadas por alcalinidade com a adição de  $200 \mu\text{l}$  de solução II ( $1\%$  SDS,  $0.2 \text{ M}$  NaOH), foi homogeneizada por agitação manual e ficou em repouso em banho de gelo por 5 minutos. Adicionaram-se  $150 \mu\text{l}$  de solução III ( $3.0 \text{ M}$  acetato de potássio, pH 5.5) foi homogeneizada por agitação manual e ficou em gelo por 5 minutos. A mistura foi centrifugada a  $14.000 \text{ g}$  por 5 minutos e o sobrenadante foi transferido para um tubo novo onde foi adicionado o mesmo volume de isopropanol  $100\%$  (aproximadamente  $400 \mu\text{l}$ ), foi misturado vigorosamente e ficou em repouso 2 minutos em temperatura ambiente. Finalmente os ácidos nucleicos foram precipitados por centrifugação a  $14.000 \text{ g}$  durante 5 minutos, o sobrenadante foi descartado e o precipitado foi lavado com a adição de  $200 \mu\text{l}$  de etanol  $70\%$ , foi feita uma última centrifugação  $14.000 \text{ g}$  durante 3 minutos e cuidadosamente foi removido o sobrenadante. Os precipitados foram colocados a  $40^\circ\text{C}$  durante 25 minutos para eliminar os resíduos de etanol. O DNA foi diluído em  $60 \mu\text{l}$  de  $\text{H}_2\text{O}$  mmQ e quantificado no espectrofotômetro NanoDrop 1000 (UNISCIENCE<sup>®</sup>), seguindo as especificações do fabricante. Foi guardado a  $-20^\circ\text{C}$  até ser usados.

#### 4.2.1.2 Construção gênica dos genes S6PP dentro do plasmídeo de expressão pET15b

A partir da extração do DNA plasmidial, os genes S6PP.1s e S6PP-2Ds foram liberados do plasmídeo pUCampMinusMCS mediante digestão com as enzimas XhoI e BamHI (New England Biolabs<sup>®</sup>). O vetor pET15b (Addgene<sup>®</sup>) foi linearizado mediante digestão enzimática usando as mesmas enzimas de restrição acima

mencionadas. A reação enzimática dupla se deu seguindo as instruções do fabricante.

Os produtos da digestão enzimática (genes e plasmídeos) foram separados da mistura por diferença de tamanho em gel de agarose 0,8% em tampão TAE 1X. As bandas de DNA foram visualizadas em transluminador e aquelas correspondentes aos genes foram cortadas do gel e colocadas em Parafilm (M<sup>®</sup>) para extrair o líquido apertando o pedaço de gel. A partir desse líquido extraído foi realizada a purificação do DNA, para 100 µL de amostra líquida foi adicionado 12 µL de NaOAc 3M, pH 5,2, 2,4 µL glicogênio 5mg ml<sup>-1</sup> e 200 µL de etanol 100%, a mistura foi incubada a -20°C por 15 minutos e centrifugada a 12.000 g durante 15 minutos a 4°C, o sobrenadante foi descartado e foram adicionados 300 µL de etanol 80%, foi centrifugado a 12.000 g por 5 minutos a 4°C, o sobrenadante foi descartado e o precipitado foi secado a temperatura ambiente, finalmente foi diluído em 20 µL de H<sub>2</sub>O mmQ. O produto purificado foi analisado mediante um novo gel de agarose 0,8% em tampão TAE 1X para corroborar sua integridade e pureza.

Os genes puros e o plasmídeo linearizado foram ligados com o kit da enzima T4 DNA Ligase (Invitrogen™, life Technologies) seguindo as instruções do fabricante. As construções receberam o nome de S6PP.1s\_pET15b e S6PP-2Ds\_pET15b.

Bactérias eletrocompetentes de *Escherichia coli* (*E. coli*) linhagem DH10B foram transformadas eletricamente com os produtos das ligações mediante eletroporador (GENE PULSER<sup>®</sup> II, BIO-RAD), seguindo as instruções do manual do fabricante. Foram obtidos múltiplos clones e os plasmídeos foram isolados mediante preparo de DNA plasmidial em pequena escala (Mini Prep) como explicado no ponto 4.2.1.1. A integridade das construções foi visualizada mediante gel de agarose 0,8% em tampão TAE 1X e foi realizada uma digestão enzimática com as enzimas *Bam*HI e *Xho*I (New England Biolabs<sup>®</sup>) para corroborar o sucesso da construção do vetor pET15b.

As construções gênicas foram corroboradas por sequenciamento realizado pela Dra. Tatiana Corrêa, especialista do laboratório GaTE usando o método de Sanger no sequenciador 3730 DNA Analyzer (Applied Biosystems).

#### 4.2.1.3 Construção gênica dos genes *S6PP* dentro do plasmídeo de expressão *pETSUMO*

Foram feitas mais duas construções gênicas usando o plasmídeo de expressão *pETSUMO* modificado a partir do plasmídeo comercial *pET28*. Os sítios de restrição *BamHI* e *Sall* foram incorporados nas sequências dos genes *S6PP*, mediante PCR usando os primers *S6PP.1s-2Ds-BamHI\_Forward* e *S6PP.1s-Sall\_Reverse* (tabela 3) para a sequência *S6PP.1s* e os primers *S6PP.1s-2D-BamHI\_Forward* e *S6PP-2D-Sall\_Reverse* (tabela 3) para a sequência *S6PP-2Ds*. Nas reações de PCR foram utilizados 0,8 µg de miniprep, Buffer 1X, 1 mM de MgCl<sub>2</sub>, 0,2 mM de dNTPs, 0,5 µM de primers e 0,2 u de enzima, numa reação de 50 µl. As condições de amplificação foram; um ciclo de desnaturação inicial a 98°C por 1 minuto, seguido de 30 ciclos de desnaturação a 98°C por 30 segundos, alinhamento dos primers a 65°C por 25s e extensão a 72°C por 2 minutos, para concluir com um ciclo de extensão final a 72°C por 7 minutos. Foi usado o kit da enzima DNA polimerase *Phusion High-Fidelity* (Thermo Scientific®) de maneira estratégica para obter extremidades tipo ponta cega (*blunt*) e facilitar a ligação ao vetor intermediário *pJET 1.2 blunt* do kit *CloneJET PCR* (ThermoFisher Scientific®). Este vetor serve para fazer a clonagem destes genes com os novos iniciadores, por ele permitir a ligação de bordas tipo *blunt* (ponta cega) de produto de PCR sem necessidade de um sítio específico. O produto da ligação no vetor *pJET 1.2 blunt* foi usado para transformar bactérias eletrocompetentes de *E. coli* linhagem DH10B mediante eletroporador (GENE PULSER® II, BIO-RAD), seguindo as instruções do manual do fabricante. Foi feito um preparo de DNA plasmidial em pequena escala (ponto 4.2.1.1). As construções no vetor intermediário *pJET 1.2 blunt* foram digeridas com as enzimas de restrição *BamHI* e *Sall* (New England Biolabs®), os genes foram purificados a partir de gel de agarose como explicado no ponto anterior (4.2.1.2). Os genes purificados e o vetor linearizado foram ligados mediante o kit da enzima T4 DNA ligase (Invitrogen™, life Technologies). As construções genicas obtidas receberam o nome de *S6PP.1s\_pETSUMO* e *S6PP-2Ds\_pETSUMO*, foram clonadas crescendo bactérias *E. coli* linhagem DH10B eletroporadas (GENE PULSER® II, BIO-RAD). Os plasmídeos foram purificados mediante preparo de DNA plasmidial em pequena escala.

As sequências das construções gênicas geradas neste ponto foram sequenciadas pela Dra. Tatiana Corrêa, especialista do laboratório GaTE usando o método de Sanger no sequenciador 3730 DNA Analyzer (Applied Biosystems) e analisadas por BLAST na plataforma NCBI (The National Center for Biotechnology Information, pelas siglas em inglês).

#### 4.2.1.4 Padronização da expressão da proteína recombinante

Para realizar os testes de expressão de proteína recombinante foram testadas diferentes combinações de variáveis, tais como plasmídeos, linhagens bacterianas de expressão, concentração de indutor do operon *lac* (isopropil- $\beta$ -D-thiogalactopiranosido (IPTG), temperatura e tempo de indução (tabela 5) a fim de encontrar a melhor condição para expressão de proteína recombinante na fração solúvel.

Tabela 5. Variáveis testadas na expressão da proteína recombinante

Plasmídeos	Construções genicas	Linhagens bacterianos	[ ] de indutor (IPTG)	Temperatura	Tempo	
pET15b	S6PP.1s_pET15b	BL21 (DE3), (Novagen®)	1 mM	37°C	3h	
		Rosetta (Novagen®)	0,7 mM	30°C		
pETSUMO	S6PP-2Ds_pET15b	C41 (Lucigen®)	0,5 mM	25°C	5h	
		S6PP.1s_pETSUMO	Tuner (DE3) (Novagen®)	0,3 mM	18°C	18h
			S6PP-2Ds_pETSUMO	CodonPlus®(DE3)-RIPL (Agilent Technologies®)		
		Arctic (Agilent Technologies®)	0,3 mM	12°C	24h	

As bactérias de expressão foram transformadas eletricamente mediante eletroporador (GENE PULSER® II, BIO-RAD), como as construções genicas mencionadas acima e inoculadas a noite inteira em meio solido LB com o antibiótico de seleção, Canamicina 50  $\mu$ l ml<sup>-1</sup> para o plasmídeo pETSUMO e Ampicilina 100  $\mu$ l ml<sup>-1</sup> para o plasmídeo pET15b mais o antibiótico de seleção da respectiva linhagem de *E. coli* seguindo as orientações do fabricante. Colônias transformadas e isoladas foram usadas para crescer pré-inoculos bacterianos de 2 ml de médio liquido LB, contendo os antibióticos de seleção. As bactérias foram incubadas a 37°C por 16 horas sob agitação de 230 RPM em shaker Innova®44 (New Brunswick scientific).

No dia seguinte, uma alíquota de 1 ml de pré-inóculo foi inoculada em 50 ml de médio LB com o antibiótico de seleção. O inóculo foi crescido a 37°C, 230 RPM em shaker Innova<sup>®</sup>44 (New Brunswick scientific) até alcançar uma densidade ótica (OD) 600 nm de  $\approx 0,8$  de absorbância. Ao atingir a absorbância desejada, 1 ml da cultura foi separada para ser usada como controle negativo não induzido, e ao resto da cultura foi adicionado IPTG. A cultura foi colocada nas diferentes temperaturas de indução durante diferentes tempos como mencionado anteriormente.

#### 4.2.1.5 *Testes de solubilidade para eleição da melhor cepa de expressão*

As células induzidas foram recuperadas por centrifugação 4.500 g por 10 minutos a 4°C e ressuspensas em 5 ml de tampão de lise (HEPES K<sup>+</sup> 100 mM, NaCl 500 mM, glicerol 10%, lisozima 0,5 mg ml<sup>-1</sup>, suplementado com inibidor de protease e fosfatase 2X, Halt<sup>™</sup> Protease & Phosphatase Inhibitor Single-Use Cocktail, EDTA-free, Thermo Scientific<sup>®</sup>, pH 7,5) as amostras permaneceram agitando em gelo durante 1 hora, passado o tempo, as células foram lisadas mediante sonicação com 30% de amplitude durante 1 minuto com ciclos de 25s *ON* e 30s *OFF*, sempre mantido em gelo. As frações solúveis e insolúveis foram separadas mediante centrifugação 15.000 RPM durante 30 min. As frações solúveis (sobrenadante) e insolúvel (precipitado) foram analisadas num gel de poliacrilamida de eletroforese SDS-PAGE (do inglês, *Sodium dodecyl sulfate-polyacrylamide gel electrophoresis*) 12% em tampão Tris-Glicine 1X (SAMBROOK et al., 1989).

#### 4.2.1.6 *Indução em larga escala para purificação da proteína recombinante*

A partir da padronização da expressão e dos testes de solubilidade conseguimos obter a melhor linhagem de *E. coli* para expressão de proteína recombinante na fração solúvel, a bactéria designada para expressar em larga escala foi CodonPlus<sup>®</sup>(DE3)-RIPL (Agilent Technologies<sup>®</sup>) transformada com as construções e com o plasmídeo pET15b. 2 L de cultura em médio líquido LB divididos em 4 frascos Erlenmeyer de 2 L foram inoculados com 10 ml de pré-inóculo preparado como explicado no ponto (4.2.1.5). Até alcançar a OD<sub>600</sub>  $\approx 0,8$  de absorbância, foi induzido com IPTG (0,3 mM) a 16°C, durante 18h aproximadamente. As células foram recuperadas por centrifugação 4.500g por 20 minutos e cada pellet celular proveniente de 1 L de cultura foi ressuspensado em 20 mL de tampão de lise filtrado (filtro 0,45  $\mu$ m), permaneceram agitando em gelo

durante 1 hora e foram lisadas com sonicação com 30% de amplitude durante 2 minutos com ciclos de 20s *ON* e 30s *OFF*, foram congeladas instantaneamente com  $Ni_2$  líquido, descongeladas e sonicadas de novo durante 2 minutos, durante todos os ciclos de sonicação os *pellets* foram mantidos em gelo. A fração solúvel (sobrenadante) e insolúvel (precipitado) foram separadas mediante centrifugação 15.000 RPM durante 45 min. A fração solúvel foi incubada com RNase A ( $1 \mu g ml^{-1}$ ) e DNase I ( $5 U ml^{-1}$ ) durante 15 minutos. Este sobrenadante foi empregado para purificar a proteína de interesse.

#### 4.2.1.7 Padronização da purificação da proteína recombinante pelo método de (IMAC)

Valendo-nos da afinidade aos metais da cauda de histidina localizada na posição N-terminal do vetor de expressão e que é expressada em conjunto com a proteína de interesse, se procedeu á purificação da proteína alvo mediante colunas contendo metal pelo método de IMAC (do inglês, *immobilized metal affinity chromatography*). Foi realizada uma busca pelo metal capaz de deixar uma proteína mas pura, como medida para otimizar a purificação, para isso foram realizados testes usando a coluna Ni-NTA Superflow Cartridge, Qiagen®, de 1 ml que é carregada com  $Ni^{2+}$  na forma de  $NiSO_4$  (50mM) e a coluna HiTrap™ IMAC FF, GE Healthcare, de 1 ml carregada com  $Co^{2+}$  na forma de  $CoCl_2$  (100 mM). A preparação da coluna foi realizada com o tampão  $A_N$  (HEPES  $K^+$  25 mM, NaCl 500 mM, glicerol 10%, imidazol 5 mM, pH 7,5) para coluna Ni-NTA e o tampão  $A_C$  (HEPES  $K^+$  25 mM, NaCl 500 mM, glicerol 10%, pH 7,5) para a coluna HiTrap™ com  $Co^{2+}$ . O carregamento do metal e limpeza da coluna foi realizado seguindo as instruções do fabricante. A amostra foi carregada na coluna através de bomba peristáltica com fluxo de  $1 ml min^{-1}$ . A purificação foi realizada em condições não desnaturantes separando as proteínas por gradiente linear de eluição com imidazol no equipamento de cromatografia líquida de alta eficiência (HPLC, do inglês *High performance liquid chromatography*) modelo ÄKTA Purifier 900 (GE Healthcare Life Science). O sistema foi programado para fazer uma lavagem inicial com 10 volumes de coluna (VC), com tampão  $A_N$  na purificação com coluna de Ni-NTA e foram eluídas mediante um gradiente linear de 0 a 100% de tampão B (HEPES  $K^+$  25 mM, NaCl 500 mM, glicerol 10%, imidazol 500 mM, pH 7,5). Já, no caso da purificação com coluna HiTrap com  $Co^{2+}$ , a lavagem inicial se deu com 10 VC de tampão  $A_C$  e o

gradiente linear de eluição se deu de 0 a 30% de tampão B. Em ambos casos o fluxo sempre foi mantido em 1 ml min<sup>-1</sup>. E a mAU (absorbância 280 nm) foi sempre monitorada e frações de 2 ml foram coletadas ao longo do processo.

As frações coletadas foram analisadas mediante géis de proteína SDS-PAGE 10% (SAMBROOK et al., 1989), e submetidas a reação de imunoblot com o anticorpo primário anti-5His (Invitrogen®) e o anticorpo secundário Anti-Mouse IgG-Alkaline Phosphatase (Sigma-Aldrich®), revelado com a solução do substrato NBT/BCIP em tampão fosfatase.

Aquelas frações mais limpas e com proteína de interesse foram concentradas utilizando concentradores Amicon Ultra-4/ultracell (Millipore) de 30 K e 100 K de acordo com as instruções do fabricante. As frações concentradas foram visualizadas em gel de proteína SDS-PAGE.

#### **4.2.2 Ensaio de atividade enzimática das duas versões de S6PP**

Os ensaios de atividade enzimática foram realizados baseados no método usado por LUNN et al., (2000), com 0,125 mM de substrato sucrose-6-fosfato (Suc6P) em 150 µl de tampão de atividade (HEPES K<sup>+</sup> 25 mM, MgCl<sub>2</sub> 5 mM, pH 7,2) foram adicionados 3 µl de extrato proteico S6PP correspondente às proteínas heterólogas S6PP.1 e S6PP-2D, a proteína BSA (albumina de soro bovino) (Sigma®) foi utilizada como controle negativo. A reação se deu a 30°C durante 30 minutos. A determinação da hidrólise mediante a quantificação de fosfatos inorgânicos (Pi) se deu usando o ensaio de fosfato inorgânico de AMES (1966) adaptado para o volume de reação determinado nos testes (150 µL).

Foi realizado um teste contemplando vários tempos de reação e duas temperaturas. A hidrólise do substrato foi medida depois de 30 min, 1, 2, 3, 4, 5 e 18 horas a 30 e 21°C a reação se deu em 150 µl de tampão. A quantificação de Pi hidrolisado foi sempre realizada mediante o método de AMES (1966).

## 5. RESULTADOS E DISCUSSÃO

### 5.1 ESTUDO MOLECULAR DO ENTORNO GENOMICO DOS GENES *S6PP*

#### 5.1.1 Genômica comparativa

Na tabela 6 encontra-se listado de forma resumida o conteúdo dos fragmentos genômicos analisados. Dezesete BACs da variedade SP80-3280 contém a duplicação em tandem dos genes *S6PP*, enquanto que na variedade R570 a duplicação foi encontrada em quatro BACs. O tamanho da região genômica em tandem é variável entre os fragmentos em estudo e vai desde 5468 pb na BAC 441\_A04 de SP803280 até 6067 pb da BAC 149\_N10 de SP803280. Vemos de igual forma que a quantidade de genes presentes na vizinhança assim como elementos de transposição é variável, tanto entre as duas variedades de cana-de-açúcar como nas BACs de uma mesma variedade. Já, as semelhanças entre fragmentos como os genes e/ou TEs compartilhados podem ser observados na figura 7.

Tabela 6. Lista dos fragmentos genômicos identificados em 21 BACs contendo a duplicação dos genes *S6PP*.

Variedades de cana-de-açúcar	Nome da BAC	Tamanho da BAC (pb)	Região genômica em tandem* (pb)	Genes vizinhos	TEs
R570	085_J04	123.896	5584	8	4
	096_D24	109.397	5586	7	2
	104_G22	99.804	5584	6	4
	237_G04	94.577	5589	5	3
SP803280	109_H09	89.338	5470	5	1
	149_N10	140.447	6067	6	3
	228_N18	141.113	5653	8	2
	273_J18	130.293	5585	9	4
	257_A23	123.970	5571	6	3
	258_B07	133.749	5585	8	4
	264_H19	118.514	5470	10	1
	264_P08	132.964	5588	12	2
	441_A04	143.827	5468	10	5
	453_B01	134.780	5571	7	4
	456_J23	126.018	5585	8	4
	465_H15	155.144	5585	9	4
	480_A02	135.325	5588	7	4
	492_H21	99.828	5585	5	3
	494_M23	135.410	5571	8	4
	524_N01	20.541	5571	1	2
	548_I17	126.419	5588	4	5

\*Esta região compreende desde o códon de início do primeiro gene até o códon de parada do gene duplicado in tandem, incluindo exons, íntrons e região intergênica.

A análise do ambiente genômico suporta que estes genes estejam localizados em um locus único com variantes genômicas estruturais compartilhadas entre os dois cultivares, como denotado pelos genes vizinhos e inserções de elementos de transposição. É interessante notar que os dois cultivares compartilham padrões de inserção dos elementos de transposição, mas também apresentam inserções que não são compartilhadas, resultando então, em que não há um padrão de inserção exclusivo para cada variedade (figura 7).

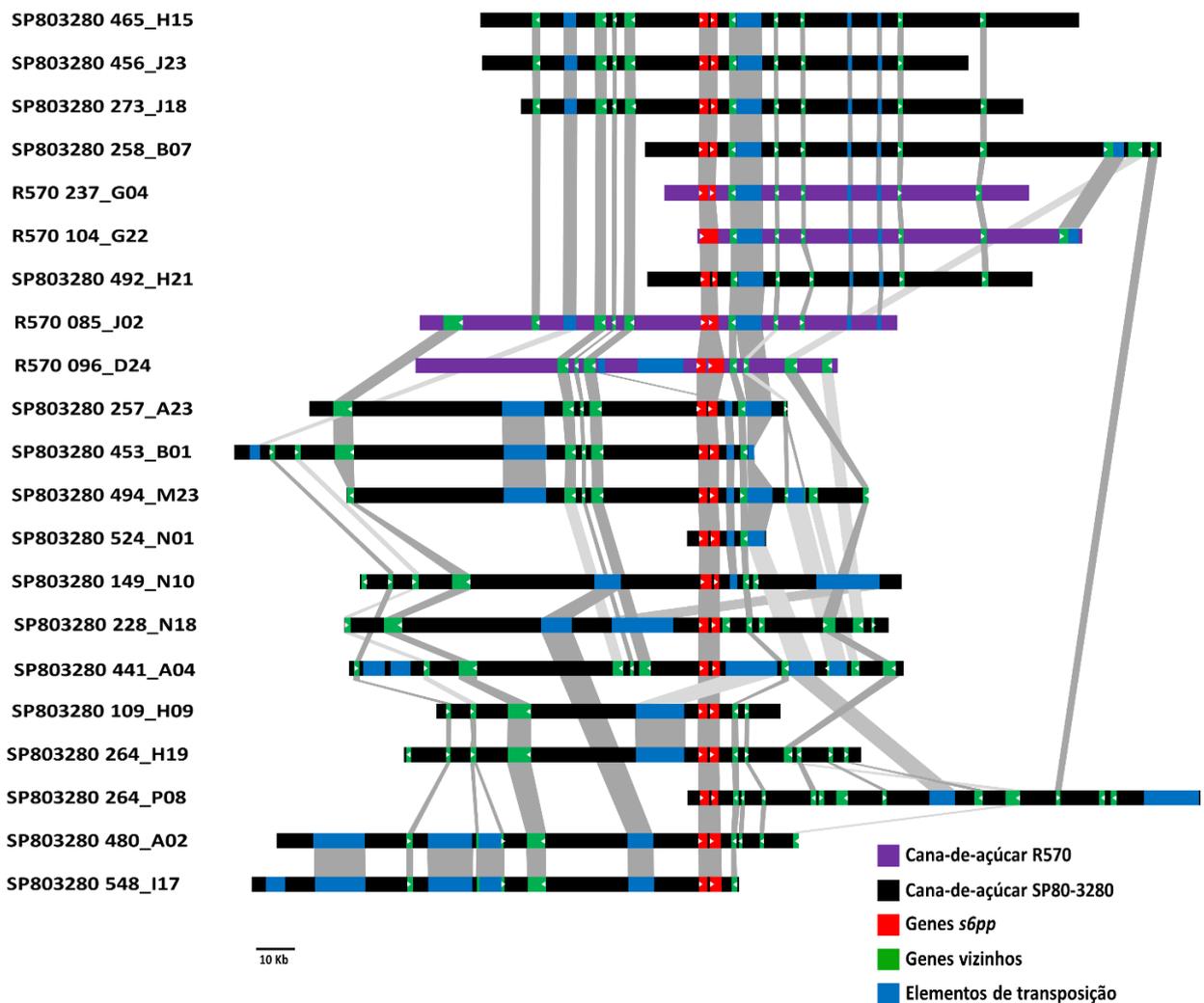


Figura 7. **Representação gráfica dos fragmentos genômicos BACs que contém as sequencias dos genes S6PP.** Retângulos pretos indicam as BACs da variedade SP80-3280; retângulos roxos representam as BACs da variedade R570, retângulos vermelhos representam os genes S6PP, retângulos verdes representam genes vizinhos de outra natureza, as setas em branco dentro dos retângulos representam o sentido da transcrição, retângulos azuis representam elementos de transposição, o sombreamento entre os retângulos indica que se trata do mesmo gene ou elemento de transposição.

Para melhorar nosso conhecimento sobre a história evolutiva destes fragmentos genômicos, foi construída uma árvore filogenética a partir do alinhamento da região genômica em tandem dos genes *S6PP* que compreende mais de 5.000 bases por BAC, começando no ATG de *S6PP.1* até o códon de parada de *S6PP.2*. A figura 8 mostra que as diferenças entre os BACs dessas cultivares híbridas não são suficientes para agrupá-las distintamente, pois os agrupamentos formados misturam as duas variedades. Observamos um grupo interessante que contém as BACs da R570 preditas a transcrever um transcrito alternativo (R570 085\_J02, R570 104\_G22) onde foram agrupados no mesmo clado BACs da SP80-3280 que podem ser interessantes para estudo.

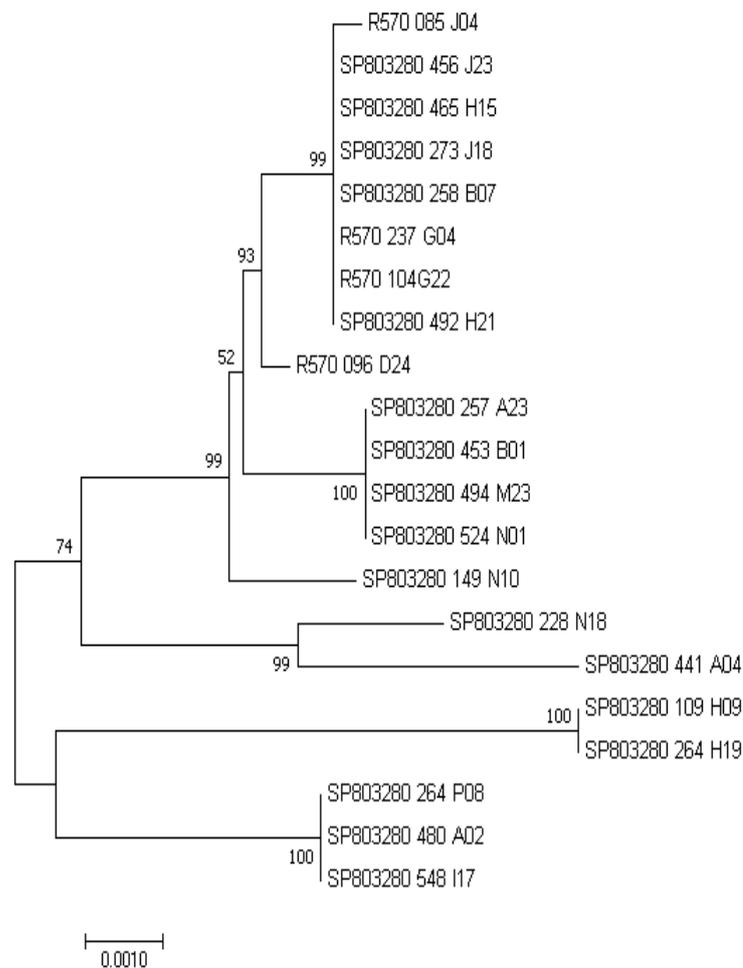


Figure 8. **Filogenia da região em tandem *S6PP* de BACs de cana-de-açúcar híbridas, R570 e SP80-3280.** Árvore filogenética *bootstrap consensus* tipo Maximum Likelihood inferido pelo software MEGA7 com o alinhamento das sequências da região em tandem. O modelo de substituição usado foi T92+G, sendo o melhor ranqueado. 1000 replicas bootstraps foram estabelecidos neste análise.

### 5.1.2 Análises filogenéticas em plantas

Numa busca exaustiva no banco de dados Phytozome12® (2018), foram encontrados outros genomas de plantas que compartilham a duplicação em tandem dos genes *S6PP*, indicando que a duplicação não é exclusiva da cana-de-açúcar, pois encontra-se presente nos genomas de *Sorghum bicolor*, *Setaria itálica*, *Setaria viridis*, *Panicum virgatum*, *Panicum hallii* e *Miscanthus sinensis*, todos eles pertencem a família das gramíneas da qual a cana-de-açúcar também é membro, porém, esta duplicação também não é exclusiva das gramíneas, pois há evidência que também está presente no genoma de *Aquilegia coerulea* que pertence à família Ranunculaceae. Embora a duplicação esteja presente em pelo menos duas famílias de plantas (dos genomas disponíveis para consulta no Phytozome12), é possível notar que a região genômica em tandem guarda diferenças em relação ao tamanho, sendo ao redor de 5000 pb na família das gramíneas e ao redor de 8000 pb nas Ranunculáceas (tabela 7). Esta duplicação em tandem não é observada nos genomas sequenciados de milho, arroz e *Brachypodium*, também pertencem a família das gramíneas.

Tabela 7. Lista de plantas que contem a duplicação em tandem dos genes *S6PP* no seu genoma (phytozome12).

Classificação por família	Genoma Phytozome12	Localização cromossômica	Nome do transcrito em Phytozome12		Região genômica em tandem* (pb)
			<i>S6PP.1</i>	<i>S6PP.2</i>	
Poaceae (Gramíneas)	<i>Sorghum bicolor</i> v3.1.1	9	Sobic.009G040900.2	Sobic.009G041000.1	5198
	<i>Sorghum bicolor</i> Rio v2.1	9	SbRio.09G043600.1	SbRio.09G043700.1	5198
	<i>Setaria itálica</i> v2.2	Scaffold_3	Seita.3G059500.2	Seita.3G059600.1	4886
	<i>Setaria viridis</i> v2.1	3	Sevir.3G060400.2	Sem anotação**	4891
	<i>Panicum virgatum</i> v1.1	3a	Pavir.Ca01201.1	Pavir.Ca01200.1	5504
	<i>Panicum virgatum</i> v1.1	Contig00021	Pavir.J00020.1	Pavir.J00021.1	6301
	<i>Panicum hallii</i> v3.0	3	Pahal.3G115900.1	Pahal.3G116000.1	5581
	<i>Panicum hallii</i> var. <i>hallii</i> v2.1	3	PhHAL.3G110000.1	PhHAL.3G110100.1	5590
	<i>Miscanthus sinensis</i> v7.1	16	Misin16G055500.1	Misin16G055600.1	5129
	<i>Miscanthus sinensis</i> v7.1	17	Misin17G0551400.1	Misin17G051300.1	5290
Ranunculácea	<i>Aquilegia coerulea</i> v3.1	1	Aqcoe1G073100.1	Aqcoe1G073000.1	8369

\*Esta região compreende desde o códon de início do primeiro gene até o códon de parada do gene duplicado em tandem, incluindo exons, íntrons e região intergênica.

\*\*Duplicação evidenciada mediante BLAST, sem anotação.

O CDS do gene S6PP.1 da variedade R570 foi usada para recuperar, através da ferramenta BLAST 2.3.0+, todos os genes disponíveis com um score maior de 150. A partir destes CDS, uma árvore filogenética foi construída com o alinhamento múltiplo da sequência de aminoácidos que correspondem aos domínios proteicos da proteína S6PP. Naqueles genomas que contem a duplicação em tandem, as sequências correspondentes as duas proteínas foram selecionadas e usadas para esta análise. De acordo a figura 9 os domínios foram agrupados pelas sequências em duplicação, vemos que aquelas correspondentes a primeira sequência caíram num mesmo grupo e aquelas pertencentes a sequência duplicada em tandem caíram em um outro grupo, além disso todas as gramíneas se agruparam num mesmo ramo. Vemos que milho (*Zea mays*) e arroz (*Oryza sativa*) que não contem a duplicação nos seus genomas, estão no ramo que separou os domínios da proteína 1, enquanto que *Oropetium thomaeum* e as duas espécies de *Brachypodium* agrupam-se no grupo dos domínios da sequência 2. As famílias restantes foram agrupadas separadamente com exceção da *Aquilegia coreoula* que contém a duplicação e cada um dos domínios ficou segregada em grupos diferentes, o domínio da proteína 1 ficou no grupo do *Theobroma cacao* (família Malvaceae), em quanto que o domínio da proteína 2, ficou no grupo da *Ananas comosus* (família Bromeliácea). O domínio da bactéria *Synechocystis* sp. (filo cianobactéria) resultou ser o mais distante.

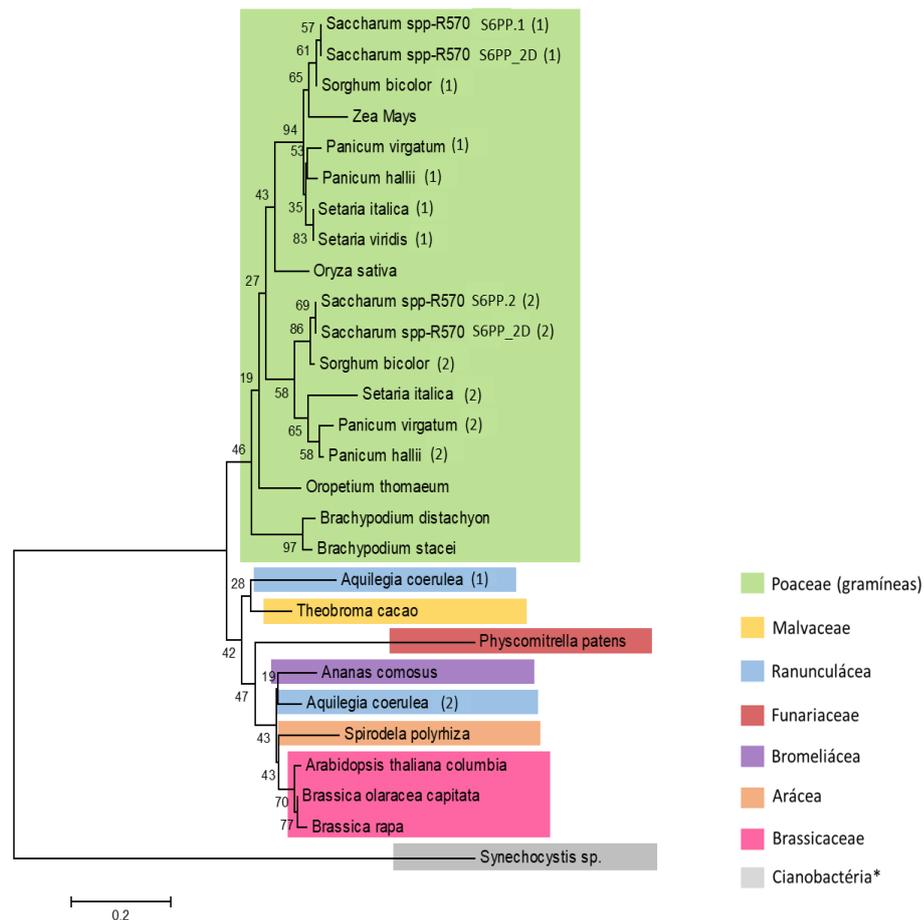


Figura 9. **Árvore filogenética dos domínios proteicos da proteína S6PP de varios genomas de plantas.** Árvore tipo neighbor-joining da filogenia dos domínios de plantas diferentes, Cada cor indica a família de plantas a qual corresponde cada genoma como descrito nos cuadros da direita. O (\*) indica que trata-se de classificação de filo de bacteria. Os nomes de genomas que terminam com numero, seja 1 ou 2, indica que esse genoma contem o gene *S6PP* duplicado em tandem, o numero 1 corresponde a primeira sequencia *S6PP* e o numero 2 indica a sequencia duplicada.

### 5.1.3 Análise de expressão em bancos de dados públicos de RNAseq

A fim de entender a expressão dos genes *S6PP* nas duas versões (*S6PP.1* e *S6PP\_2D*) (figura 5) e com o objetivo de encontrar um transcrito para nossa proposta de nova isoforma do gene *S6PP* (versão *S6PP\_2D*) provavelmente decorrente de *splicing* alternativo, foi realizado um mapeamento em bancos de dados de RNAseq gerados a partir de bibliotecas de cana-de-açúcar de diferentes idades e submetidas

a diferentes condições de desenvolvimento, como explicado no ponto 4.1.2 da metodologia e resumido na seguinte tabela 8. O mapeamento dos CDSs preditos de cada gene contra as leituras geradas em estudo e depositado no NCBI foi realizado com o programa bowtie2 disponibilizado no portal do CEFAP/USP. Os parâmetros para a realização do alinhamento foram os definidos pelo programa como **padrão**. Os resultados do mapeamento foram normalizados por FPKM (para sequenciamento *paired end*) e RPKM (para sequenciamento *single end*), a partir da cobertura dos *reads* frente ao CDS em análise. A sequência codificante do gene sacarose fosfato sintase (*SPS*) foi usado como controle da expressão por codificar uma proteína que atua no passo prévio a reação catalisada pela S6PP, na via da síntese de sacarose e no armazenamento da sacarose nos vacúolos.

Tabela 8. Bancos de dados de RNAseq gerado a partir do sequenciamento de diferentes bibliotecas de cana-de-açúcar sometidas em diferentes condições de desenvolvimento.

Autores	Cultivar	Tecidos	Idade	Infeção	Nome das amostras
GaTE Lab -IB/USP (não publicado)	R570	Meristema	3 meses	Não infetada	Meristema
		Raiz		Não infetada	Raiz
SCHAKER et al., (2016) (SRA SRX884131)	RB92-5345	Brotos	5 dias*	Não infetado (controle)	5-DC
				<i>Sporisorium scitamineum</i>	5-DAI
		Base do chicote	200 dias*	Não infetado (controle)	200-DC
				<i>Sporisorium scitamineum</i>	200-DAI
SANTA BRIGIDA et al., (2016) (SRA SRP041671)	SP70-1143	Planta inteira	7 dias**	Não infetado (controle)	Mock
				<i>Acidovorax avenae</i> subsp. <i>avenae</i>	Infected

O (\*) indica que a idade das plantas coletadas corresponde a 5 e 200 dias depois da infecção de plantas que foram infetadas no momento da plantação. O (\*\*) indica que a idade quando foram coletadas as plantas corresponde a 7 dias depois da infecção, já, a infecção se deu na planta no momento do enraizamento de pois da plantação.

### 5.1.3.1 Meristema e raiz

O resultado obtido a partir do banco de banco de dados do Laboratório GaTE (dados não publicados), gerado a partir de amostras de cana-de-açúcar R570 com 3 meses de idade, foi a nossa principal motivação, resultando no mapeamento com uma cobertura total de *reads* no transcrito da isoforma proposta neste trabalho, a versão *S6PP\_2D* que sugere uma conjugação das duas sequencias em tandem, decorrente de um provável evento de intronização da região intergênica das duas sequencias e uma exonização de uma pequena porção, que dá lugar a um transcrito que expressa um peptídeo com dois domínios da *S6PP*, encontrado na raiz deste experimento. Vemos também, que os genes *S6PP* e *SPS* são majoritariamente expressos na raiz quando comparado com o meristema apical (figura 10), os transcritos dos genes usados nos mapeamentos fazem parte da mesma via metabólica na síntese de sacarose e no preparo da sacarose prévio a entrada no vacúolo onde é acumulada. A raiz e o meristema são tecidos não fotossintetizantes, incapazes de sintetizar sacarose, então, a expressão desses genes, sugere que participam no acúmulo da sacarose e corroboram HATCH; SMITH, (1984), que evidenciam a expressão *S6PP* em raiz de cenoura.

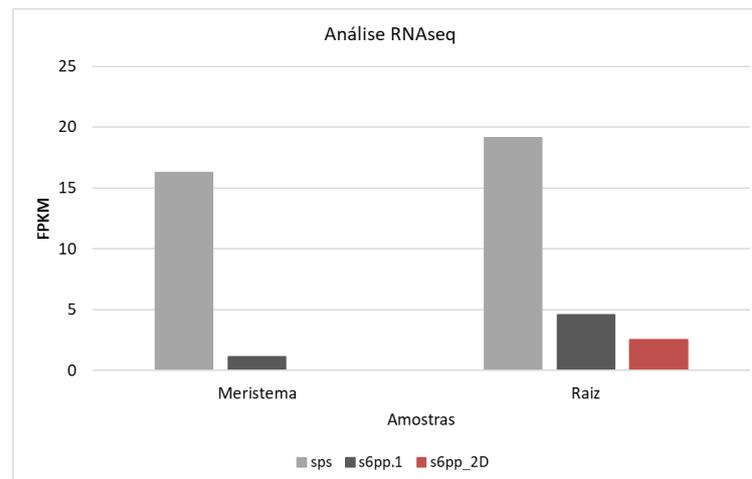


Figura 10. **Gráfico de barras dos valores de FPKM determinado a partir do análise da expressão por RNAseq de amostras de cana-de-açúcar R570.** Raiz, amostra com 3 meses de idade; Meristema, amostras com 3 meses de idade. Barra cinza claro, FPKM do gene *SPS*; barra cinza escura, FPKM do gene *S6PP.1*; barra vermelha, FPKM do gene *S6PP\_2D*.

### 5.1.3.2 Inoculação com *Sporisorium scitamineum*

SCHAKER et al., (2016) investigaram por RNAseq a interação do fungo *S. scitamineum*, causador do carvão, em cana de açúcar. Amostras da variedade RB92-5345 foram coletadas de brotos com 5 dias depois da inoculação (5-DAI) e na base do chicote até 2 cm abaixo do colmo com 200 dias depois da inoculação (200-DAI) e seus respectivos controles (não inoculado) com 5 dias (5-DC) e 200 (200-DC) para gerar as bibliotecas de RNAseq. O estudo foi desenvolvido com 3 réplicas biológicas e o desvio padrão foi calculado baseado no resultado do mapeamento por separado de cada replica. Na figura 11 podemos observar que não há evidência da expressão do transcrito *S6PP\_2D* nos tecidos analisados e que a expressão do transcrito *S6PP* é afetada pela presença do patógeno em 200DAI. As amostras com 5 dias, tanto inoculadas como não inoculadas, a expressão do gene *S6PP* se mantém perto do mesmo valor de FPKM, nas amostras controle (não infetadas) de 200 dias (200-DC) o valor da expressão aumenta um pouco quando comparado com as amostras de 5 dias, já, nas amostras infetadas com 200 dias (200-DAI) o valor do FPKM cai por abaixo do valor das amostras com 5 dias. Mesmo padrão, foi encontrado no gene *SPS*. Segundo os autores SCHAKER et al., (2016), 200 dias é um estágio de infecção quando a doença causada pelo fungo é evidente e de alguma forma isso compromete o metabolismo da sacarose, pois com a infecção vemos uma diminuição na expressão de dois genes que participam neste metabolismo.

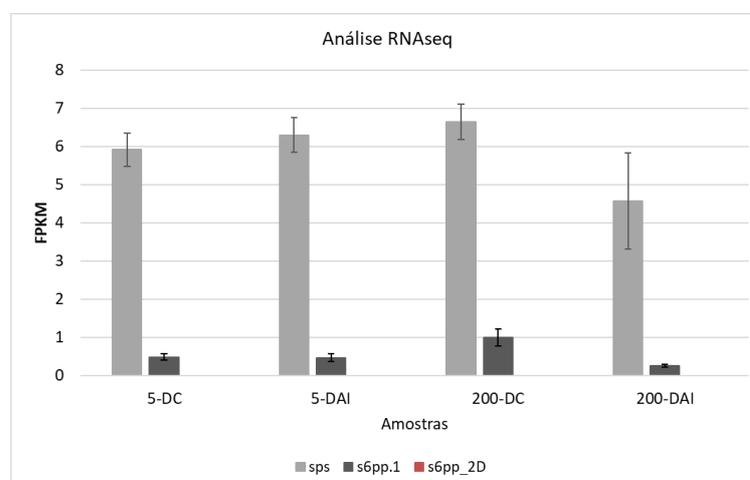


Figura 11. Gráfico de barras dos valores de FPKM determinado a partir da análise da expressão por RNAseq de amostras de cana-de-açúcar SP70-1143. 5-DAI, brotos coletados com 5 dias depois da inoculação; 5-DC, brotos de cinco dias não inoculados; 200-DAI, base do chicote até 2 cm abaixo do colmo com 200 dias depois da inoculação com *Sporisorium scitamineum*; 200-DC, base do chicote até 2 cm abaixo do colmo com 200 dias de crescimento não inoculado. Barra cinza claro, FPKM do gene *SPS*; barra cinza escuro, FPKM do gene *S6PP.1*.

### 5.1.3.3 Inoculação com *Acidovorax avenae* subsp. *avenae*

SANTA BRIGIDA et al., (2016) inocularam canas-de-açúcar da variedade SP70-1143 já enraizadas, com a bactéria *Acidovorax avenae* subsp. *avenae*, (causante da doença da falsa estria vermelha), após de 7 sete dias da inoculação (infected), as plantas inteiras foram coletadas para extração do RNA e preparo das bibliotecas para o sequenciamento, o experimento foi realizado com duplicatas biológicas e o controle livre do patógeno também foi estudado (Mock).

Baseados nos resultados do cálculo do RPKM o transcrito do gene *S6PP* na isoforma *S6PP\_2D* não foi encontrado. O transcrito *SPS* é majoritariamente expresso em amostras controle (sem infecção) com um valor de RPKM de 27 e nas amostras infetadas resultou em RPKM aproximadamente 15. Já, o gene *S6PP.1* apresenta um resultado inverso, com aumento da expressão nas amostras infetadas (RPKM aprox. 12) enquanto que, nos controles apresenta RPKM similar ao controle de aproximadamente 10. O que sugere que a expressão deste gene não foi comprometida pela infecção bacteriana. Porém, os valores do desvio padrão são muito discrepantes neste experimento (figura 12).

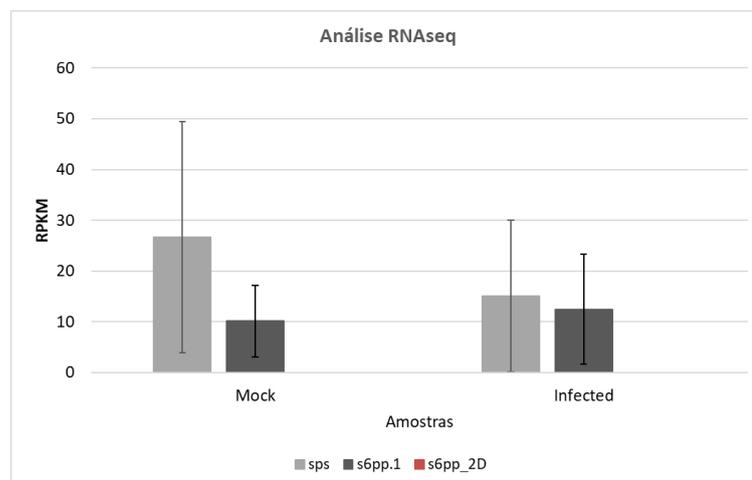


Figura 12. **Gráfico de barras dos valores de RPKM determinado a partir do análise da expressão por RNaseq de amostras de cana-de-açúcar RB92-5345.** Infected, planta inteira infetada com *Acidovorax avenae* subsp. *avenae* após o enraizamento e coletada 7 dias após a inoculação; Mock, planta inteira usada como controle, sem infecção. Barra cinza claro, RPKM do gene *SPS*; barra cinza escura, RPKM do gene *S6PP.1*.

#### 5.1.4 Perfil da expressão dos genes *S6PP* em diferentes tecidos de cana-de-açúcar ao longo do desenvolvimento.

A fim de entender o padrão de expressão que seguem os genes *S6PP* nas duas versões (*S6PP.1* e *S6PP\_2D*) foi montado um experimento para quantificação da expressão em 4 variedades de cana de cana-de-açúcar, híbridas, R570 e SP80-3280 e as espécies *Saccharum officinarum* (variedade Badilla) e *S. spontaneum* (variedade IN84-58). O desenho experimental compreende a coleta de diferentes tecidos (folha, meristema apical, colmo e raiz) em três estádios de desenvolvimento (3, 6 e 9 meses). A seleção das amostras é decorrente da análise de integridade do RNA total extraído mediante visualização por gel de agarose.

O desenho dos iniciadores para qPCR se deu envolvendo uma junção exon-exon com a intenção de reduzir o risco de amplificação de DNA genômico e evitar falsos positivos, os iniciadores amplificam regiões de 100 pb em regiões estratégicas como ilustrado na figura 6 na seção das metodologias.

##### 5.1.4.1 Busca do melhor gene normalizador para quantificação da expressão por qPCR

A fim de definir o melhor normalizador para os estudos de PCR quantitativa, foram testados os iniciadores que amplificam regiões codificantes dos genes *Act*, *GAPDH* e *PUB*, sobre as amostras da variedade híbrida R570.

De acordo a figura 13, podemos ver a diferença de expressão entre os três genes para os quais testamos os pares de iniciadores, os Ct de expressão do gene *GAPDH* varia de 22 (palmito de 6 meses) até 27 (raiz de 3 meses), os Ct do gene *Act* varia de 23,5 (palmito 6 meses) até ≈36.5 (palmito 3 meses), enquanto que os Ct do gene *PUB* varia de 23 (palmito 9 meses) até 27,5 (colmo 6 meses). É fácil notar que *PUB* resultou na menor variabilidade de expressão quando comparada com *Act* e *GAPDH*, e foi, portanto, o gene normalizador usado na quantificação das amostras usadas neste trabalho. ISKANDAR et al., (2004) suporta que o gene *Act* e *GAPDH* é muito variável nos diferentes tecidos de cana-de-açúcar cultivar Q117, não sendo bons normalizadores da expressão genica em análise qPCR.

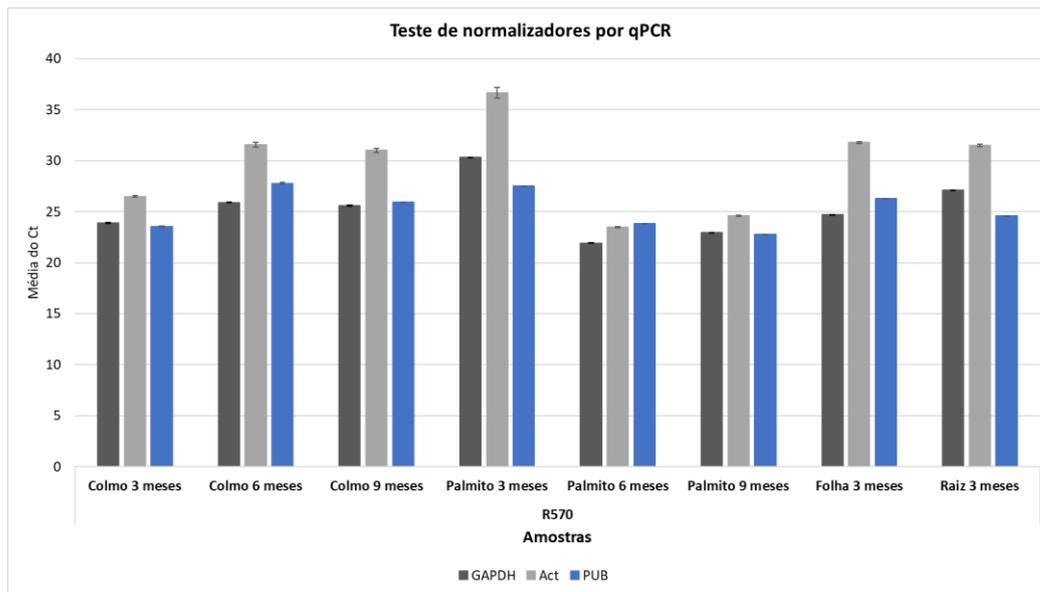


Figura 13. **Gráfico de barras do teste de genes normalizadores para qPCR.** No eixo vertical é mostrado a média do Ct, em quanto que, no eixo horizontal é apresentado as amostras de tecidos da variedade R570 de cana-de-açúcar de diferentes idades. Barras cinza escura representa o Ct do gene *GAPDH*; Barras cinza claro, representa o Ct do gene *Act*; Barras azuis, representa o Ct do gene *PUB*.

#### 5.1.4.2 Análise quantitativa da expressão dos genes *S6PP*

Um experimento para abordar o padrão de expressão desses genes em cana-de-açúcar foi preparado com a seguinte justificativa: a sacarose é produzida pela primeira vez na folha como resultado da fotossíntese; sendo então translocado através do floema (colmo) para os tecidos dreno (meristema e raízes). Para verificar a transcrição em cana-de-açúcar das duas isoformas, *S6PP.1* e *S6PP.2D*, examinamos o padrão de expressão de duas variedades híbridas SP80320 e R570, bem como duas espécies progenitoras *S. officinarum* (Badilla) e *S. spontaneum* (IN8458).

Os resultados são apresentados na figura 14 e confirmam que a isoforma *S6PP.2D* é expressa principalmente em três meses na maioria dos tecidos examinados, também a isoforma convencional *S6PP*, porém a expressão não é uniforme ao longo do desenvolvimento em ambos os casos. Ambas as variedades comerciais têm níveis de expressão semelhantes nas folhas, no entanto, *S. spontaneum* e *S. officinarum* têm expressão contrastante. Curiosamente, cada variedade se comporta de maneira diferente quanto aos tecidos dreno, meristema e

raízes. A zona de transporte, representada pelo colmo, é similar entre as duas variedades comerciais e Badilla, enquanto a expressão em IN84-58 é elevada. O tecido radicular de Badilla tem a maior expressão dos dois transcritos em todos os tecidos examinados.

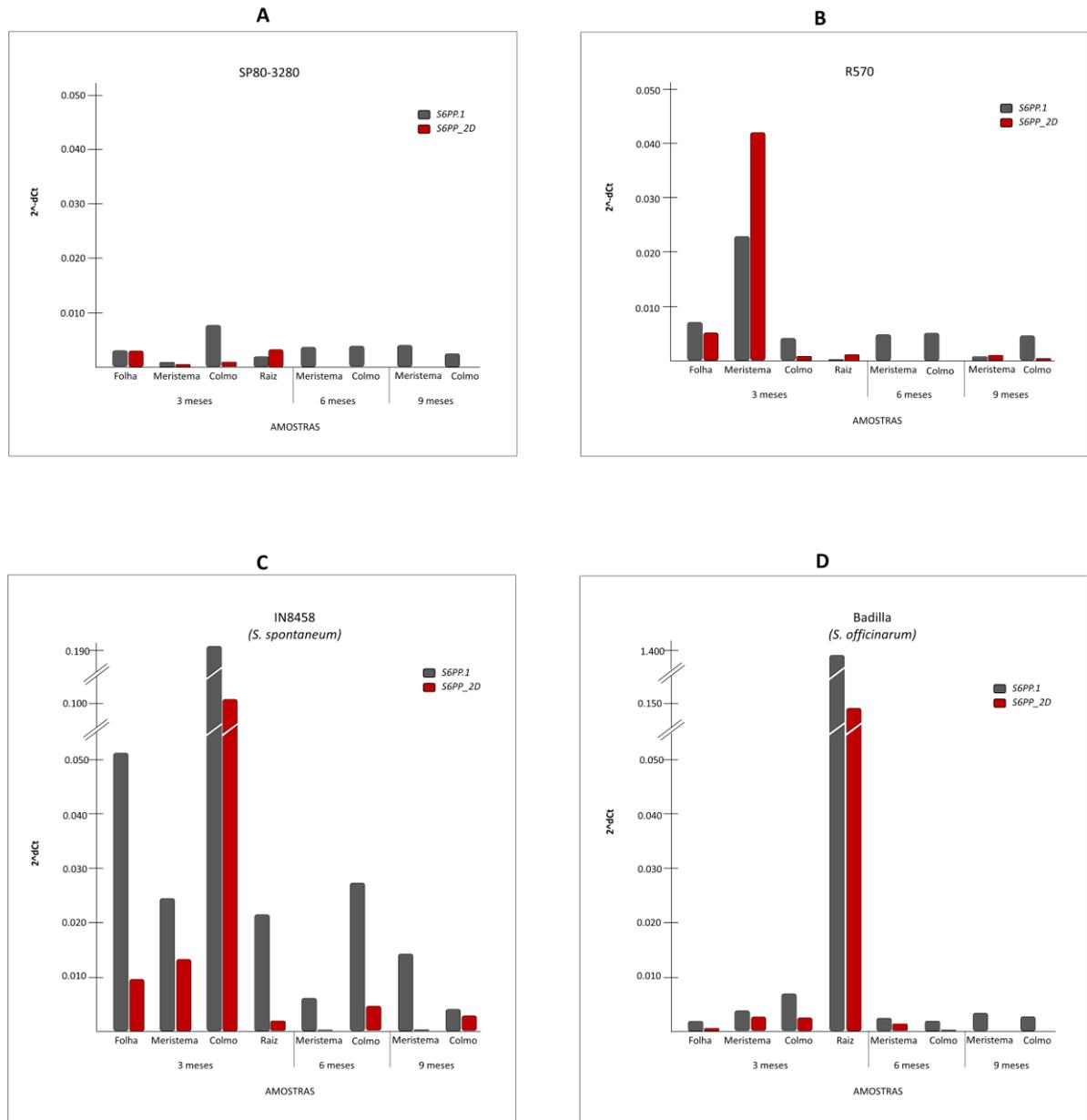


Figura 14. Gráfico de barras da análise de expressão dos genes *S6PP* ao longo do tempo (3, 6 e 9 meses) em folha, meristema, colmo e raiz. A) Padrão de expressão da variedade híbrida SP80-3280. B) Padrão de expressão da variedade híbrida R570. C) Padrão de expressão da espécie *S. spontaneum* (variedade IN8458). D) Padrão de expressão da espécie *S. officinarum* (variedade Badilla). As barras cinza escuro representam a expressão do gene da versão *S6PP.1*, as barras vermelhas representam a expressão do gene da versão *S6PP.2D*. No eixo das abscissas (x) são as amostras ao longo do tempo. No eixo das ordenadas (Y) está o nível de expressão determinado pelo  $2^{-dCt}$ .

A presença da enzima S6PP é esperada em folhas, um tecido fotossintetizante capaz de produzir sacarose a partir da molécula Suc6P, substrato da S6PP na via metabólica da síntese de sacarose. Esta sacarose livre é translocada através dos vasos condutores do floema para todos os tecidos não fotossintéticos, incluindo tecidos dreno, onde é acumulada após sofrer uma série de reações de interconversão, ativação para Suc6P e desfosforilação pela S6PP antes da entrada no vacúolo onde é armazenada na forma de hexoses (glicose e frutose). A exportação de carboidratos das folhas fotossintetizantes (fontes) fornece substrato para o crescimento e manutenção de tecidos vegetais não fotossintéticos (AINSWORTH; BUSH, 2010). Os açúcares representam a principal fonte de energia para todos os organismos eucarióticos, os carboidratos são essenciais para processos fundamentais no crescimento das plantas. A cana de açúcar acumula mais sacarose perto da última fase do ciclo de desenvolvimento, quando tem redução na taxa de crescimento. Mas, e os níveis de expressão?

De acordo com o experimento apresentado acima, observamos maior expressão de ambos transcritos nos tecidos com 3 meses de idade e uma diminuição notável com o passar do tempo. Aos 3 meses, a cana de açúcar está em fase de crescimento exigindo maiores quantidades de sacarose para realizar as funções vitais da planta. A sacarose que está sendo produzida durante a fotossíntese nas folhas, translocado para os órgãos heterotróficos e armazenado, é rapidamente utilizado pelo metabolismo respiratório para produção de energia necessária aos processos de divisão celular, diferenciação e crescimento vegetal (WANG et al., 2013; RUAN 2014). O que faz com que as enzimas envolvidas na síntese da sacarose, tanto na fotossíntese quanto no armazenamento, estejam em atividade constante (evidenciada pela expressão dos transcritos) permitindo o equilíbrio cíclico entre síntese, armazenamento, uso e nova síntese (figura 15-A). Portanto, o transporte e partição de açúcares de folhas fototróficas (fonte) para órgãos heterotróficos (dreno) através do floema são os principais parâmetros que controlam a produtividade das culturas (AINSWORTH; BUSH, 2010).

Em contraste, como ilustrado na figura 15-B, a cana-de-açúcar tende a acumular sacarose quando tem redução na taxa de crescimento (ROBERTO, 2015), quando atinge a idade madura, entre 10 e 14 meses dependendo da variedade (HAGOS; MENGISTU; MEQUANINT et al. 2014). A atividade catalítica da enzima S6PP é

comprometida na presença de sacarose, trata-se, portanto, de uma inibição pelo produto da reação. 50 mM de sacarose reduz em 60% a atividade da S6PP parcialmente purificada a partir da variedade Pindar de cana-de-açúcar (HAWKER; HATCH, 1966). S6PP de cana-de-açúcar (*Saccharum* spp.cv NCO310) e beterraba vermelha foram inibidos 40% e 59%, respectivamente, na presença de sacarose 100 mM (HAWKER et al. 1987). S6PP de arroz (*Oryza sativa*) apresentou 15% de inibição em 160 mM de sacarose, (LUNN et al. 2000). S6PP de *Synechocystis* sp. O PCC 6803 é inibido em 19% e 27% na presença de sacarose 200 mM e 660 mM, respectivamente (LUNN, 2002). Cabe notar que CARSON et al. (2002), relatam transcrição diferencial de genes em colmos maduros e imaturos, revelando que algumas das enzimas envolvidas no metabolismo da sacarose apresentam expressão mais abundante em tecidos imaturos, ao mesmo tempo em que revelaram que os genes que apresentaram maiores níveis de transcrição em tecidos maduros não estão envolvidos no metabolismo da sacarose. Pode-se concluir que a modulação da expressão gênica e da atividade enzimática de S6PP se dá por um controle de feedback

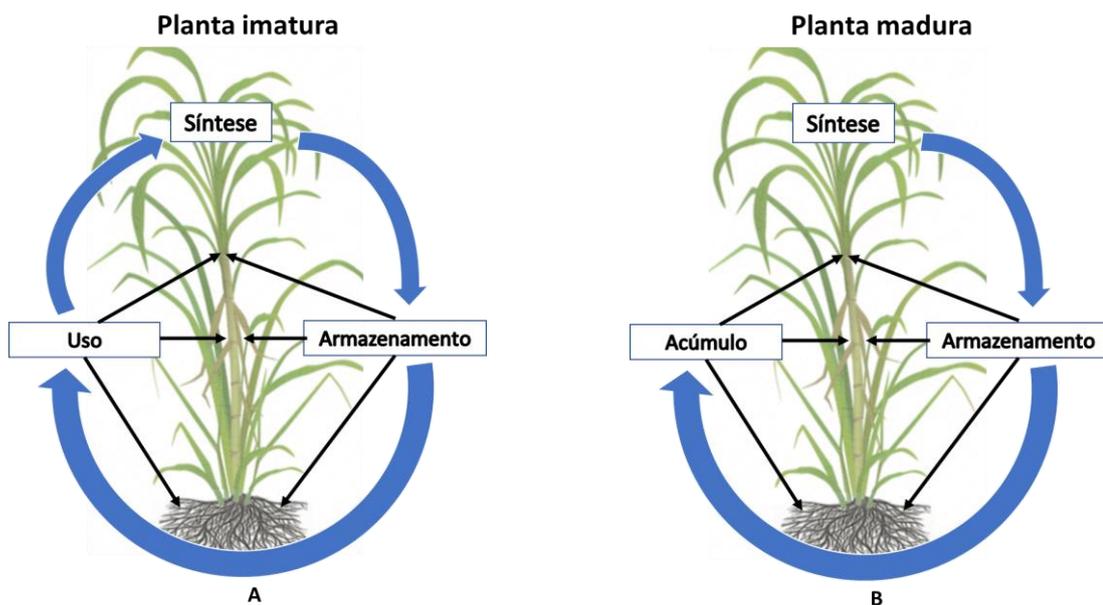


Figura 15. **Modelo do ciclo da sacarose durante o desenvolvimento da planta em idades jovens (planta imatura) e idades velhas (planta madura).** A) Ciclo da sacarose na etapa imatura. A sacarose é sintetizada nas folhas, depois de transportada é armazenada nos tecidos não fotossintetizantes e é usada para desenvolver as funções vitais durante o crescimento. B) Ciclo da sacarose na etapa madura. A sacarose sintetizada nas folhas é transportada aos tecidos não fotossintetizantes, é armazenada e começa a ser acumulada porque a taxa de crescimento diminui.

O estudo apresentado nesta dissertação também mostrou diferenças de expressão entre os tecidos. Segundo WANG et al., (2013), o acúmulo de sacarose depende do tamanho e atividade do tecido. As diferenças de expressão entre variedades são dadas pela capacidade da cultura de acumular sacarose. A espécie *S. spontaneum* (variedade IN-8458), em geral, apresentou a maior expressão das enzimas S6PP (S6PP.1 e S6PP\_2D), nos tecidos e idades analisadas, esta espécie é caracterizada por não acumular grandes quantidades de sacarose, no entanto, é uma espécie com alto nível de resistência a doenças, adaptabilidade e tolerância ao estresse. Por outro lado, sabe-se que *S. officinarum* (variedade Badilla), que apresentou níveis mais baixos de expressão, acumula altos níveis de sacarose na célula, mas tem baixa resistência a doenças. Através do cruzamento seletivo destas duas espécies surgem híbridos interespecíficos SP80-3280 e R570, cuja identidade genômica corresponde a 70-80% de *S. officinarum* e 10-23% de *S. spontaneum*, dependendo da cultivar (HEINZ 1991; D' HONT; GLASZMANN, 2001). Neste estudo, os dois híbridos apresentaram baixos níveis de expressão, semelhantes a *S. officinarum*. WANG et al. (2013) declaram que variedades comerciais híbridas são capazes de acumular altas concentrações de sacarose, até 650 mM ou 18% de seu peso fresco. Então, postulamos que a inibição da atividade da enzima S6PP esteja relacionada ao acúmulo de sacarose nos tecidos nas fases tardias do desenvolvimento.

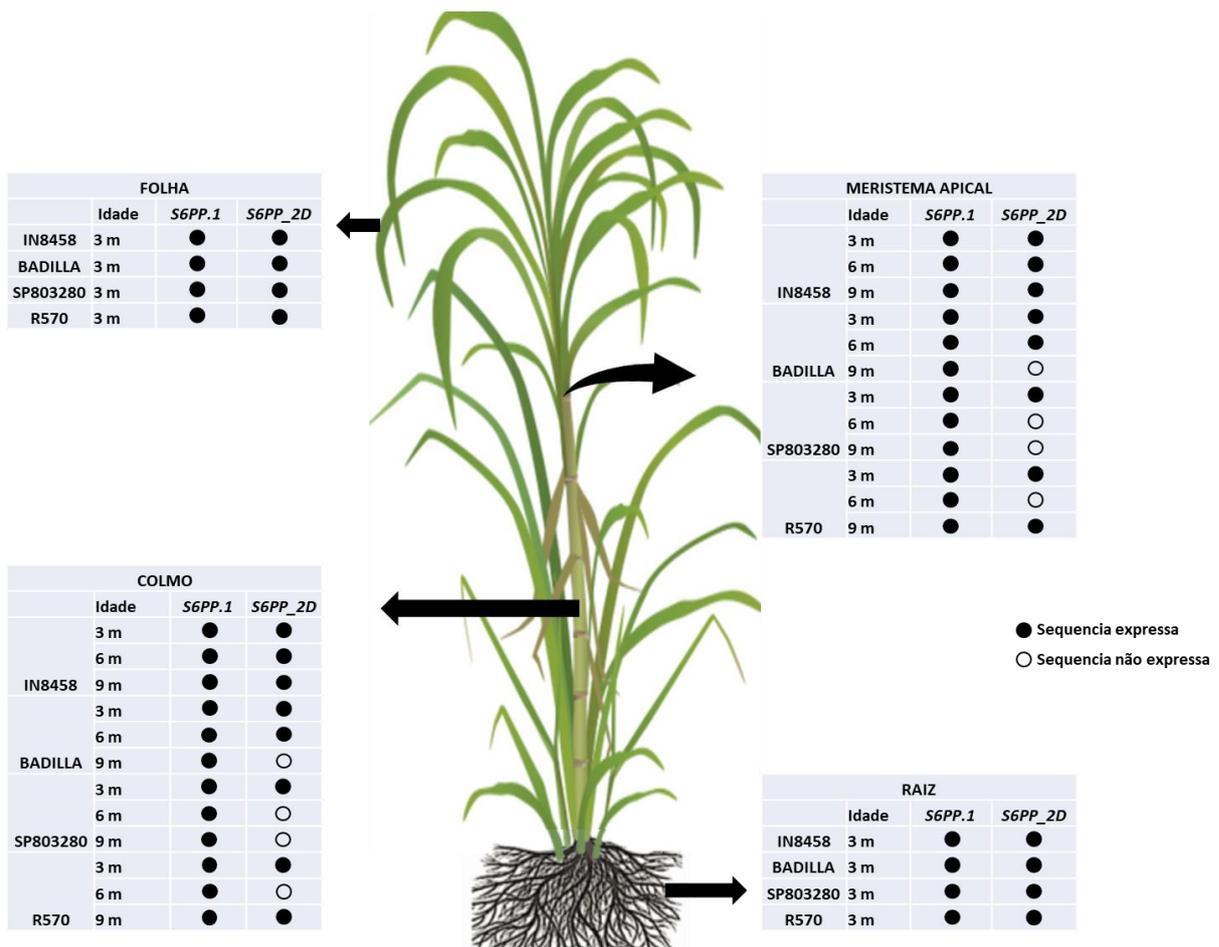


Figura 16. **Mapa biológico da expressão dos genes *S6PP.1* e *S6PP\_2D* nos diferentes tecidos analisados.** O círculo preenchido em preto indica as condições em que houve presença do transcrito, o círculo em branco representa a ausência do transcrito. A idade da planta é apresentada em meses.

A Figura 16 resume o padrão de presença/ausência do transcrito *S6PP\_2D* apoiando que o evento de exonização deve vir dos genomas de ambas as espécies parentais que deram origem às variedades comerciais híbridas. O transcrito *S6PP\_2D* seguiu um padrão de presença e ausência de expressão no colmo e meristema, *S6PP\_2D* foi expressa em IN8458 nos três estádios de desenvolvimento, enquanto que, em Badilla a presença do transcrito em 3 e 6 meses VER FINAL é detectada. As cultivares comerciais exibem padrão semelhante aos três meses, variando aos seis e nove meses. A expressão está ausente em R570 aos 6 meses e em SP803280 o transcrito é detectado apenas aos três meses. Os resultados sugerem que o gene *S6PP\_2D* de alguma forma participa do metabolismo de sacarose antes da entrada no vacúolo, principalmente nas primeiras idades de desenvolvimento da planta, neste caso, 3 meses foi a primeira idade analisada.

## 5.2 ESTUDO BIOQUÍMICO E FUNCIONAL DAS PROTEÍNAS S6PP EXPRESSADAS EM SISTEMAS HETERÓLOGOS

### 5.2.1 Padronização da expressão da proteína recombinante e método de purificação

Mediante a padronização da expressão da proteína recombinante em linhagens de *E. coli*, foi possível determinar que a melhor linhagem capaz de expressar maior quantidade de proteína alvo na fração solúvel, quando comparada com as demais linhagens testadas (exibido no ponto 4.2.1.4). A padronização e os testes de solubilidade serviram para determinar as melhores condições de expressão, resultando ser aqueles parâmetros que fazem a bactéria ter um metabolismo devagar, esses parâmetros encontram-se resumidos na tabela 9.

Tabela 9. Variáveis selecionadas após a padronização da expressão, para indução em larga escala e método de purificação.

Plasmídeo	Linhagem bacteriano	[ ] de indutor (IPTG)	Temperatura	Tempo de indução	Coluna de purificação
pET15b	CodonPlus®(DE3)-RIPL (Agilent Technologies©)	0,3 mM	16°C	18 horas	HiTrap™ IMAC FF (CoCl <sub>2</sub> )

### 5.2.2 Indução em larga escala e purificação da proteína recombinante

Após a indução da expressão das proteínas recombinantes em larga escala, foi realizada a purificação da proteína alvo a partir do extrato protéico obtido sob condições não desnaturantes. Foram usadas as mesmas condições de cultivo, indução da expressão e purificação da proteína recombinante nas duas versões da *S6PP* expostas ao longo deste trabalho. Recapitulando que *S6PP.1* é a proteína que expressa a partir de uma sequência que se encontra em tandem e que é compartilhada com outros genomas de plantas. *S6PP\_2D*, é uma isoforma proposta desta sequência que fusiona a duplicação em tandem, provavelmente decorrente de *splicing* alternativo, de modo que a isoforma resultante expresse uma proteína que contém dois domínios proteicos da *S6PP*.

### 5.2.3 Purificação da enzima S6PP.1

Durante o processo de purificação por afinidade ao  $\text{Co}^{2+}$  foram obtidos os cromatogramas (figura 17) que ilustra o resultado da purificação da proteína recombinante S6PP.1, onde, há a presença de dois grandes picos que indicam conteúdo de proteínas eluídas em diferentes concentrações de eluente (imidazol). Como exposto, todas as frações foram analisadas e todas aquelas que apresentavam o mesmo padrão de bandas foram combinadas. Dois conjuntos de frações concentradas foram obtidos, representados pelos quadros da figura 17. Estes foram manipulados por separado.

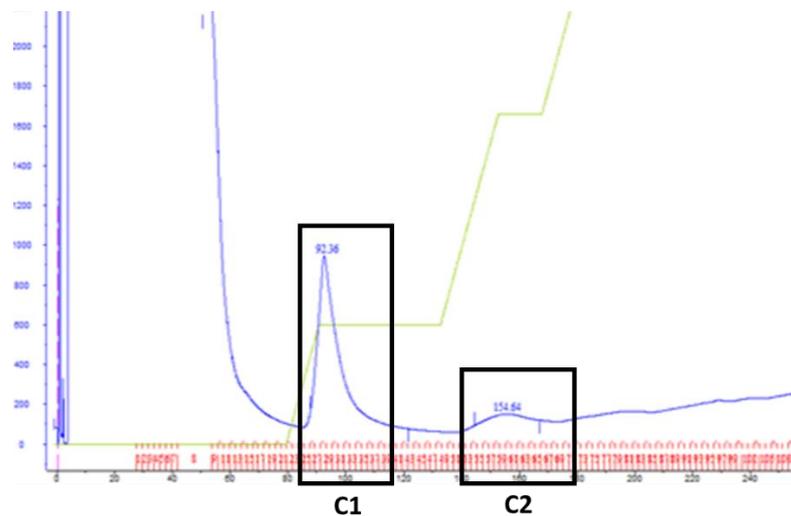


Figura 17. **Cromatograma da purificação da proteína S6PP.1.** A proteína S6PP.1 clonada no vetor de expressão pET15b, foi purificada pelo método de IMAC usando  $\text{CoCl}_2$ . O Retângulo preto C1 indica um conjunto de frações que foram concentradas; O retângulo preto chamado de C2, indica o segundo conjunto de frações concentradas.

Os conjuntos de frações foram concentrados primeiramente com concentrador de 30 kDa (Amicon Ultra-4/ultracell, Millipore). Em seguida, a fração retida foi passada por um concentrador com filtro de 100 KDa da mesma marca, e as frações retidas e não retidas foram coletadas e analisadas em SDS-PAGE. Na figura 18-A, vemos as frações concentradas do conjunto 1 (C1) e do conjunto 2 (C2), que foram concentradas em filtro de 30 KDa, é fácil notar que o conjunto mais limpo é o C2. Esses conjuntos concentrados passaram pelo filtro de 100 kDa, as frações retidas ficaram ainda mais concentradas, já, a fração atravessada, resultou numa banda limpa e totalmente isolada, nos dois conjuntos (figura 18-B).

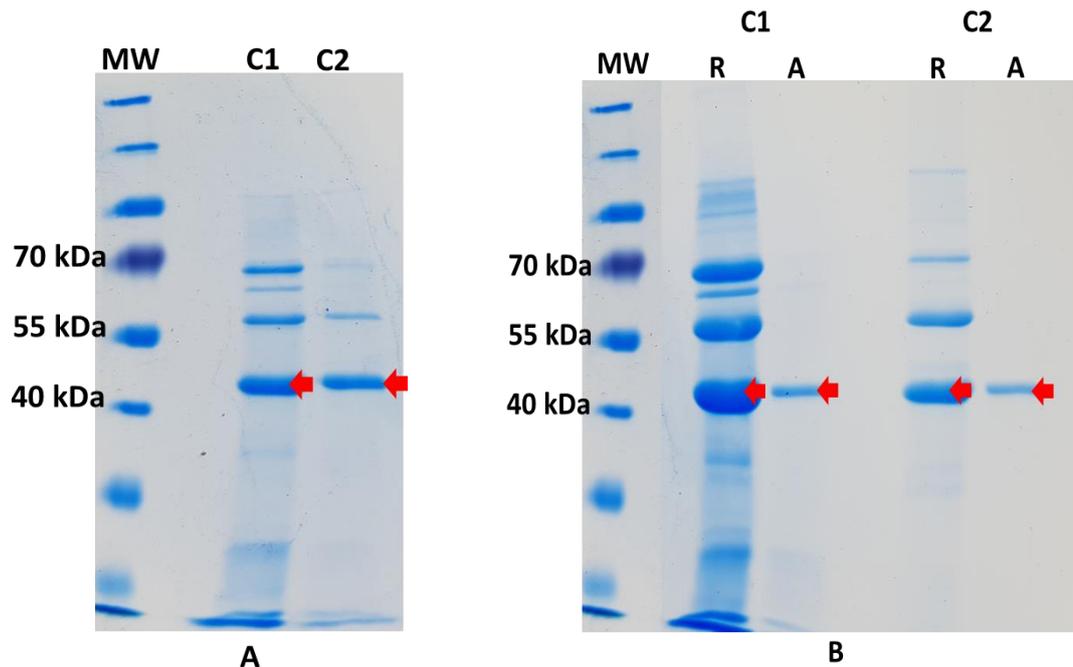


Figura 18. **SDS-PAGE dos conjuntos das frações da purificação da proteína S6PP.1\_pET15b concentradas com filtro de 30 kDa e 100 kDa.** A) Gel das frações concentradas com filtro de 30 kDa após purificação. B) Gel das frações concentradas com filtro de 100 kDa a partir da fração retida no filtro de 30 kDa. MW, indica o marcador de peso molecular; C1, conjunto 1 de frações concentradas (figura 17); C2, conjunto 2 de frações concentradas (figura 17); R, fração retida da preparação; A, fração atravessada da preparação; as setas vermelhas indicam as bandas pertencentes à proteína S6PP.1.

Fazendo uso da ferramenta bioinformática ProtParam de ExpASY (ARTIMO et al., 2012), a proteína S6PP.1 expressa pelo vetor pET15b é predita a transcrever um peptídeo de 50,6 kDa que inclui a região da cauda de histidina do plasmídeo, o que é corroborado pelo tamanho da banda do SDS-PAGE. Porém, o que chamou a atenção foi que no gel, nas colunas da fração retida concentrada com filtro de 100 kDa ficou uma banda grossa que coincide com o tamanho predito da S6PP.1 mediante ExpASY, o que sugere que a proteína não conseguiu atravessar eficazmente o filtro. Surgindo a hipótese de que a proteína possa ser um homodímero de aproximadamente 101 kDa na forma nativa e o que é visualizado no SDS-PAGE são as subunidades desnaturadas. A ideia é suportada pelos estudos prévios em outras plantas nas que é caracterizada como dímero, exemplo disso são, ervilha (*Pisum sativum*) 120 kDa (WHITAKER, 1984), arroz (*Oryza sativa*) 100 kDa (LUNN et al., 2000), *Arabidopsis thaliana* 90 kDa (ALBI et al., 2016), corroborando a hipótese de que em cana-de-açúcar possa ser também uma enzima de duas

subunidades. As frações retidas e não retidas dos dois conjuntos foram usados nos testes de atividade enzimática.

#### 5.2.4 Purificação da enzima S6PP\_2D

Já, no caso da purificação da enzima S6PP\_2D, vemos no cromatograma (figura 19) um pico bem menor, em comparação com a purificação da proteína S6PP.1. Todas as frações foram analisadas por SDS-PAGE e aquelas que concordavam com o tamanho da S6PP\_2D foram concentradas em concentrador de 50 kDa.

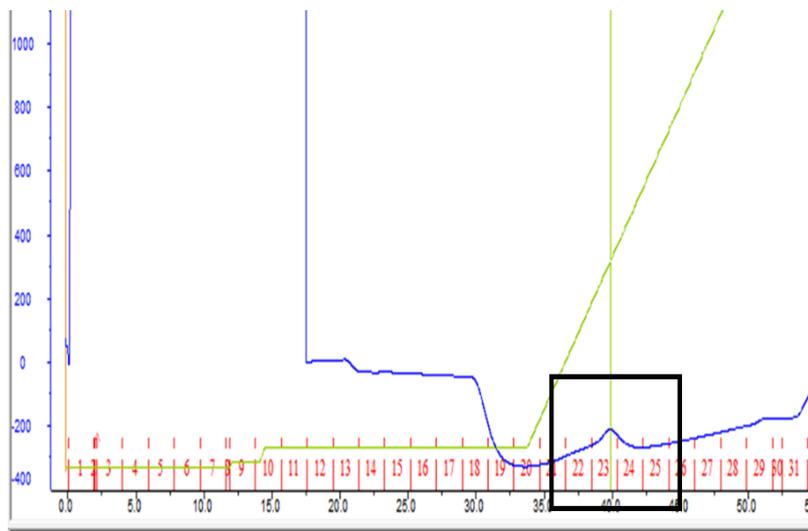


Figura 19. **Cromatograma da purificação da proteína S6PP\_2D.** A proteína S6PP\_2D clonada no vetor de expressão pET15b, foi purificada pelo método de IMAC usando  $\text{CoCl}_2$ . O retângulo preto indica o conjunto de frações que foram concentradas com filtro de 50 kDa.

A predição do tamanho da S6PP\_2D mediante ProtParam de ExPASy (ARTIMO et al., 2012) resultou ser de 90.6 kDa, que inclui a cauda de histidina do vetor pET15b que expressa em conjunto com a proteína alvo. No gel SDS-PAGE foi possível ver uma banda próximo a esse tamanho e que não está presente no extrato proteico da bactéria não induzida, que foi usada como controle negativo (não mostrado) e confirmado mediante *Western blot* (não mostrado). Como podemos ver na figura 20, as frações concentradas provenientes da purificação contêm muitas proteínas não específicas e pouca proteína alvo.

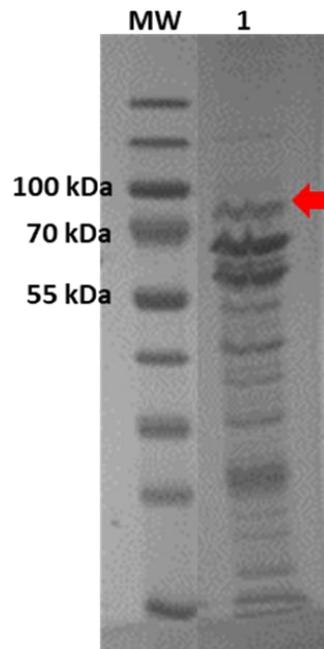


Figura 20. **SDS-PAGE** do conjunto de frações concentradas provenientes da purificação da proteína S6PP-2D\_pET15b. MW, indica o marcador de peso molecular. 1 indica o extrato enzimático proveniente do conjunto de frações concentradas com filtro de 50 kDa a partir da purificação. A seta vermelha indica a banda correspondente a proteína alvo recombinante.

O extrato semi-puro contendo a proteína recombinante S6PP\_2D obtido a partir da purificação e concentração de frações foi usado para realizar teste enzimáticos.

### 5.2.5 Ensaio de atividade enzimática das versões de S6PP

Três principais ensaios de atividade enzimática serão expostos a seguir. De acordo com os trabalhos publicados por HAWKER et al., (1987), HAWKER; HATCH, (1966), ECHEVERRIA; SALERNO, (1994), LUNN et al., (2000), a enzima fosfatase S6PP (EC 3.1.3.24) é uma enzima com alta afinidade pelo substrato sacarose 6-fosfato (Suc6P) e é possível realizar ensaios enzimáticos a partir de um extrato proteico. Por isso, se procedeu a realização de testes de atividade enzimática a partir do extrato semi-puro contendo a proteína S6PP\_2D e aqueles conjuntos concentrados contendo a proteína S6PP.1 e com a proteína S6PP.1 totalmente pura. As reações se deram em tampão HEPES K<sup>+</sup> 25mM, pH 7,2, 30°C em presença do cofator Mg<sup>2+</sup> 5mM para o qual é altamente dependente (LUNN et al., 2000).

### 5.2.6 Ensaio enzimático com extratos proteicos semi-puros

O primeiro teste consistiu em um ensaio com extratos proteicos semi-puros (não apresentados). A proteína S6PP.1 foi capaz de hidrolisar o substrato Suc6P em 60 minutos, como evidenciado pela coloração azul (AMES, 1966), a proteína S6PP\_2D não mostrou coloração positiva de fosfohidrólise nessas condições (figura 21).

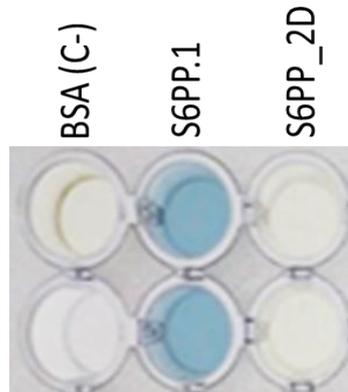


Figura 21. **Atividade enzimática das proteínas heterólogas S6PP.1 e S6PP\_2D expressadas no vetor pET15b.** De esquerda à direita; BSA, usado com controle negativo, reação enzimática da S6PP.1, reação enzimática da S6PP\_2D.

### 5.2.7 Ensaio enzimático com extratos proteicos semi-puros com variação de tempo e temperatura de reação

Outro ensaio com variação de tempo e temperatura de reação foi desenvolvido, também com extratos enzimáticos semi-puros. A hidrólise do substrato foi medida por absorbância 820 nm depois de 30 min, 1, 2, 3, 4, 5 e 18 horas de reação a 30 e 21°C. Tomando como referência o método de AMES (1966), cada 0,260 de Abs resulta em 0,01  $\mu\text{mol}$  de fosfato inorgânico (Pi), podendo, então, quantificar a hidrólise. A enzima S6PP.1, evidenciou ter atividade desde os primeiros 30 minutos de reação, nesse tempo foram hidrolisados 0,0293  $\mu\text{mol}$  de Pi a 30°C e 0,0302  $\mu\text{mol}$  de Pi a 21°C (figura 22-A). O ensaio teve resultados promissores na atividade da enzima com dois domínios catalíticos (S6PP\_2D), mostrando atividade a partir de três horas de reação, onde, de acordo as medidas de absorbância, aproximadamente foram hidrolisados 0,0010  $\mu\text{mol}$  de Pi a 21°C e 0,0014  $\mu\text{mol}$  de Pi a 30°C (figura 22-B).

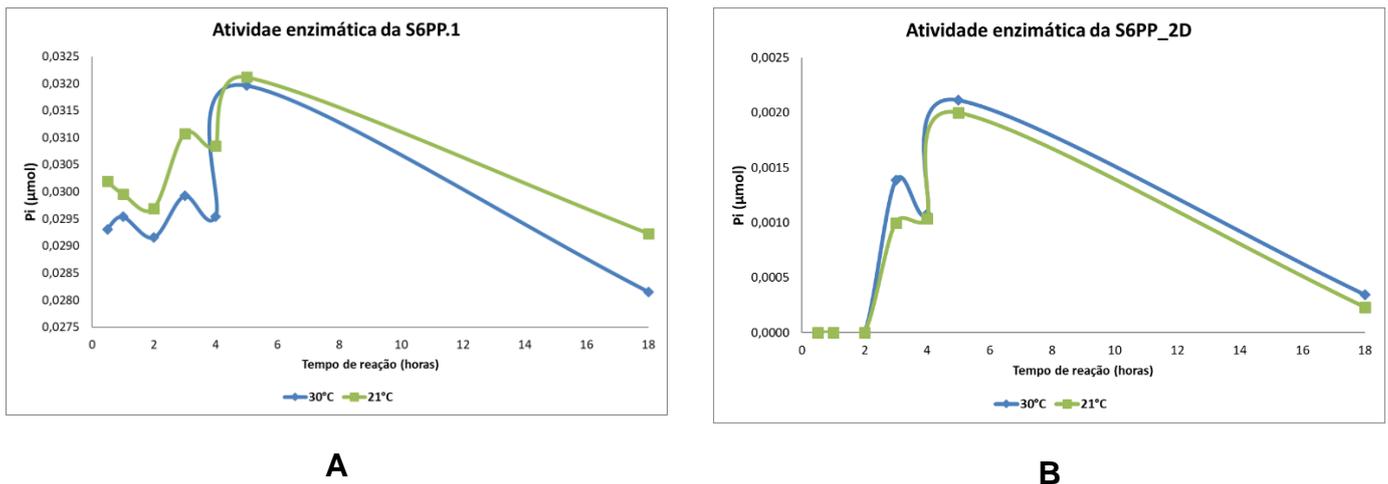


Figura 22. **Gráficos de linha do ensaio enzimático com variação de tempo e temperatura.** A) Gráfico da reação enzimática com S6PP.1. B) Gráfico da reação enzimática com S6PP\_2D. Os gráficos são apresentados mostrando o tempo de reação (30 min, 1, 2, 3, 4, 5 e 18 horas) no eixo X e o Pi hidrolisado em µmol no eixo Y. As linhas azuis representam a reação a 30°C e as linhas verdes a reação a 22°C.

Embora a atividade da S6PP\_2D seja superada 300X pela S6PP.1, é interessante notar que o comportamento da curva de reação manteve o padrão enzimático da enzima S6PP.1 ao longo do tempo. Em cinco horas de reação se detectou a maior atividade nas duas temperaturas testadas e nas duas enzimas. Porém, cinco horas foi o último tempo estipulado para fazer medição de substrato neste teste, antes de passar para atividade *over night*, onde observa-se o decaimento da atividade também para as duas proteínas. O primeiro teste realizado não mostrou atividade alguma da enzima S6PP\_2D, pois o tempo de reação mínimo que precisa para ter atividade é de três horas sob essas condições testadas. Provavelmente, a baixa atividade catalítica da enzima S6PP\_2D pode ser devida às condições de reação que não foram as adequadas para o seu máximo desempenho na atividade enzimática. BUT et al. (2013) publicaram a existência de uma enzima bifuncional sacarose fosfato sintase e ao mesmo tempo sacarose fosfato fosfatase que resulta numa molécula homotetramérica de 340 kDa encontrada na bactéria *Methylobacillus flagellatus* KT. O trabalho evidencia que esta enzima com sítio catalítico da S6PP, requer condições de reação diferentes das que são relatadas em outros estudos. O tetrâmero tem melhor atividade catalítica com uma concentração ótima de cofator  $Mg^{2+}$  17,5 mM, a melhor temperatura de reação é 45°C e o pH ideal é 6.5. Baseados neste artigo e nos resultados do ensaio obtido, vemos que a enzima S6PP\_2D mostrou maior atividade a 30°C e enzima S6PP.1 a 21°C, deixando evidências que

independente das duas conterem o mesmo sítio catalítico, há possibilidade de padronização de condições diferentes para uma catalise ótima.

### 5.2.8 Ensaio enzimático com enzima S6PP.1 pura e frações semi-puras

Esse ensaio da atividade enzimática se deu usando a proteína S6PP.1 totalmente pura e os conjuntos semi-puros que não atravessaram pelo filtro de 100 kDa, como explicado no ponto da purificação da enzima S6PP.1 (5.2.3). Foram realizados 2 tipos de controles negativos, um colocando enzima e água e outro colocando substrato e água.

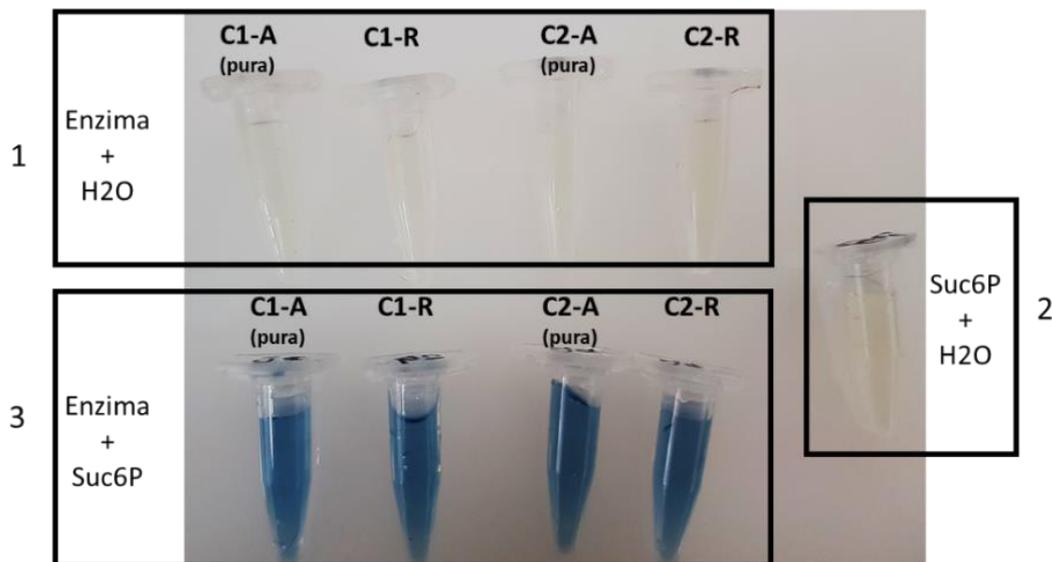


Figura 23. **Ensaio enzimático da proteína S6PP.1 pura e com extratos enzimáticos semi-puros.** O quadro 1 mostra o controle negativo usando enzima ou extrato enzimático e água (sem substrato). O quadro 2 mostra o controle negativo usando substrato e água (sem enzima ou extrato enzimático). O quadro 3 mostra a reação da enzima com o substrato. As amostras enzimáticas utilizadas foram; C1-A, fração atravessada do conjunto 1 que corresponde a enzima isolada; C1-R, fração retida do conjunto 1, que corresponde a extrato semi-puro; C2-A, fração atravessada do conjunto 2 que corresponde a enzima isolada; C2-R, fração retida do conjunto 2 que corresponde a extrato semi-puro

Neste teste todas as amostras apresentaram-se ativas conforme coloração azul intensa que resulta da presença de Pi (figura 23). O ensaio evidenciou que proteína S6PP.1 ficou na fração retida do concentrador de 100 kDa, evidenciando que a proteína na forma nativa possa ser uma molécula de mais que 50,6 kDa, provavelmente o enovelamento resulta em uma enzima com duas subunidades proteicas, como no caso da S6PP da ervilha, arroz e *Arabidopsis thaliana*.

## 6 CONCLUSÕES

Variedades híbridas comerciais de *Saccharum* spp. R570 e SP80-3280 apresentam uma duplicação em tandem dos genes S6PP. Essa duplicação em tandem não é exclusiva de cana-de-açúcar. Esta duplicação é compartilhada com outros genomas de plantas, tais como *Sorghum bicolor*, *Setaria italica*, *Setaria viridis*, *Panicum virgatum*, *Panicum hallii*, *Miscanthus sinensis* e *Aquilegia coerulea*. Evidenciando que a duplicação está presente em pelo menos duas famílias dos genomas sequenciados disponíveis, Poaceae e Ranunculácea.

A duplicação em Poaceae é ancestral ao grupo visto que os domínios S6PP se agrupam segundo as cópias S6PP (1) e S6PP (2). É interessante notar que os domínios de arroz e milho foram agrupados nos clados dos domínios da sequência 1 podendo ter perdido a duplicação do gene no transcurso da evolução enquanto *Brachypodium* e *Oropetium* não agrupam.

*Saccharum* spp. variedades híbridas R570 e SP80-3280 e as espécies ancestrais *S. officinarum* (variedade badilla) e *S. spontaneum* (variedade IN8458) apresentam uma nova isoforma deste gene S6PP, nunca antes evidenciada nem publicada, onde a região intergênica da duplicação em tandem é intronizada e uma parte é exonizada, provavelmente decorrente de *splicing alternativo*. Isto resulta em um transcrito que contém os dois domínios das duas sequências (S6PP\_2D), o peptídeo resultante, então, mantém dois domínios funcionais da S6PP.

O transcrito S6PP\_2D e S6PP são expressos em tecidos fotossintetizantes (folhas) e em tecidos não fotossintetizantes (colmo, meristema apical e raiz), evidenciando a participação na via da síntese de sacarose durante a fotossíntese e na liberação da sacarose no preparo para o armazenamento nos vacúolos dos tecidos dreno. São majoritariamente expressos em idades imaturas do desenvolvimento, em torno de 3 meses de idades, enquanto que, a expressão é diminuída em idades tardias do desenvolvimento, quando a planta alcança uma fase de maturidade, dos 6 aos 9 meses de idade.

Sabe-se que a atividade enzimática de S6PP se dá por um controle de feedback, em que na presença de sacarose acumulada a enzima é inibida repercutindo na expressão, então, postulamos que a redução da transcrição dos genes S6PP esteja

relacionada também ao acúmulo de sacarose nos tecidos nas fases tardias do desenvolvimento e evidenciaria a modulação da expressão gênica também pelo produto da reação.

As variedades híbridas apresentam menor expressão dos transcritos *S6PP*, atribuído a capacidade de acumular maior quantidade de sacarose em comparação com as espécies ancestrais.

A análise de RNAseq mostrou que a expressão do gene *S6PP* é variável sob condições de infecção, e de idade.

As enzimas recombinantes *S6PP.1* e *S6PP\_2D* tem atividade enzimática para hidrólise de fosfato inorgânico. Porém, as condições de reação enzimática da enzima com dois domínios (*S6PP\_2D*) tem a ser padronizados para melhorar o desempenho na catalise.

Este trabalho contribuiu no entendimento de uma nova isoforma do gene *S6PP* presente em variedades de cana-de-açúcar. Espera-se como resultado futuro poder agregar valor ao melhoramento genético da planta, aumentando a competitividade na produção de sacarose nas cultivares de *Saccharum* spp.

## REFERENCIAS

- AFGAN, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. **Nucleic acids research**, v. 46, n. W1, p. W537-W544, 2016.
- AGROSTAT (2016). Estatísticas de Comércio Exterior do Agronegócio Brasileiro. Site: <http://indicadores.agricultura.gov.br/agrostat/index.htm>, accessed: September 10<sup>th</sup>, 2018.
- AINSWORTH, E. A.; BUSH, D. R. Carbohydrate export from the leaf-A highly regulated process and target to enhance photosynthesis and productivity. **Plant physiology**, p. pp. 110.167684, 2010.
- ALBI, T. et al. Characterization of the Sucrose Phosphate Phosphatase (SPP) Isoforms from *Arabidopsis thaliana* and Role of the S6PPc Domain in Dimerization. **PloS one**, v. 11, n. 11, p. e0166308, 2016.
- AMES, B. N. [10] Assay of inorganic phosphate, total phosphate and phosphatases. In: **Methods in enzymology**. Academic Press, 1966. p. 115-118.
- AMES JR, L. L. Some zeolite equilibria with alkaline earth metal cations. **American Mineralogist: Journal of Earth and Planetary Materials**, v. 49, n. 7-8, p. 1099-1110, 1964.
- ARTIMO, P. et al. ExpASy: SIB bioinformatics resource portal. **Nucleic acids research**, v. 40, n. W1, p. W597-W603, 2012.
- BUT, S. Y. et al. Bifunctional sucrose phosphate synthase/phosphatase is involved in the sucrose biosynthesis by *Methylobacillus flagellatus* KT. **FEMS microbiology letters**, v. 347, n. 1, p. 43-51, 2013.
- CARSON, D. L. et al. Differential gene expression in sugarcane leaf and internodal tissues of varying maturity. **South African journal of botany**, v. 68, n. 4, p. 434-442, 2002.
- CHRISPEELS, M. J.; SADAVA, D. E. **Plants, genes, and crop biotechnology**. Jones & Bartlett Learning, 2003.
- CONAB-Companhia Nacional de Abastecimento, (2018). Acompanhamento da safra brasileira de cana-de-açúcar 2018/2019. –v.1 –Brasília. Site: <http://www.conab.gov.br>, accessed: September 10<sup>th</sup>, 2018.
- CUMINO, A.; EKEROTH, C.; SALERNO, G. L. Sucrose-phosphate phosphatase from *Anabaena* sp. strain PCC 7120: isolation of the protein and gene revealed significant structural differences from the higher-plant enzyme. **Planta**, v. 214, n. 2, p. 250-256, 2001.
- D'HONT, A.; GLASZMANN, J. C. Sugarcane genome analysis with molecular markers, a first decade of research. In: **Proc Int Soc Sugarcane Technol**. 2001. p. 556-559.

D'HONT, A. et al. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. **Genome**, v. 41, n. 2, p. 221-225, 1998.

DANIELS, J.; ROACH, B. T. Taxonomy and evolution. In 'Sugarcane improvement through breeding'.(Ed. DJ Heinz) pp. 7–84. 1987.

DE SETTA, N. et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. **BMC genomics**, v. 15, n. 1, p. 540, 2014.

DINARDO-MIRANDA, L. L.; DE VASCONCELLOS, A. C. M.; DE ANDRADE LANDELL, M. G. (Ed.). **Cana-de-açúcar**. Campinas: Instituto Agronômico, 2008.

ECHEVERRÍA, E.; SALERNO, G. Properties of sucrose-phosphate phosphatase from rice (*Oryza sativa*) leaves. **Plant Science**, v. 96, n. 1-2, p. 15-19, 1994.

FERREIRA, R. R. et al. Tolerância diferencial de variedades de cana-de-açúcar a estresse por herbicidas. **Bragantia**, v. 69, n. 2, p. 395-404, 2010.

FIEULAINÉ, S. et al. The structure of a cyanobacterial sucrose-phosphatase reveals the sugar tongs that release free sucrose in the cell. **The Plant Cell**, v. 17, n. 7, p. 2049-2058, 2005.

FIEULAINÉ, S.; LUNN, J. E.; FERRER, J.-L. Crystal structure of a cyanobacterial sucrose-phosphatase in complex with glucose-containing disaccharides. **Proteins: Structure, Function, and Bioinformatics**, v. 68, n. 3, p. 796-801, 2007.

OKURA, V. et al. A BAC library of the SP80-3280 sugarcane variety (*saccharum* sp.) and its inferred microsynteny with the sorghum genome. **BMC research notes**, v. 5, n. 1, p. 185, 2012.

Food and Agriculture Organization of the United Nations, Commodities by Country (2018); Site: [www.fao.org/faostat/en/#rankings/commodities\\_by\\_country](http://www.fao.org/faostat/en/#rankings/commodities_by_country). <http://www.fao.org/faostat/en/#data/QC>, accessed: September 10<sup>th</sup>, 2018.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current Opinion in Plant Biology**, v. 5, n. 2, p. 122-127, 2002.

HAGOS, H.; MENGISTU, L.; MEQUANINT, Y. Determining optimum harvest age of sugarcane varieties on the newly establishing sugar project in the tropical areas of tendaho, Ethiopia. **Advances in Crop Science and Technology**, 2014.

HATCH, M. D. Sugar accumulation by sugar-cane storage tissue: the role of sucrose phosphate. **Biochemical Journal**, v. 93, n. 3, p. 521, 1964.

HAWKER, J. S.; HATCH, M. D. A specific sucrose phosphatase from plant tissues. **Biochemical Journal**, v. 99, n. 1, p. 102, 1966.

HAWKER, J. S. et al. Sucrose phosphatase associated with vacuole preparations from red beet, sugar beet, and immature sugarcane stem. **Plant physiology**, v. 84, n. 4, p. 1281-1285, 1987.

HEINZ, D. J. Sugarcane cytogenetics. **Chromosome engineering in plants: Genetics, Breeding, Evolution, Part B. Elsevier, Amsterdam**, p. 279-293, 1991.

ISKANDAR, H. M. et al. Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. **Plant Molecular Biology Reporter**, v. 22, n. 4, p. 325-337, 2004.

JIANG, Shu-Ye et al. Sucrose metabolism gene families and their biological functions. **Scientific reports**, v. 5, p. 17583, 2015.

KALT-TORRES, W. et al. Diurnal changes in maize leaf photosynthesis: I. Carbon exchange rate, assimilate export rate, and enzyme activities. **Plant Physiology**, v. 83, n. 2, p. 283-288, 1987.

KITTS, D. D. Sucrose: From Field to Table. **Carbohydrate news**, 2010.

KUMAR, M.; TRIPATHI, A. To Study of the Different Modes of Tillage for the Performance of Sugarcane Cutter Planter. 2015.

LANGMEAD, B. Aligning short sequencing reads with Bowtie. **Current protocols in bioinformatics**, v. 32, n. 1, p. 11.7. 1-11.7. 14, 2010.

LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome biology**, v. 10, n. 3, p. R25, 2009.

LEMOINE, Remi et al. Source-to-sink transport of sugar and regulation by environmental factors. **Frontiers in plant science**, v. 4, p. 272, 2013.

LORIMER, G. H., BUCHANAN, B. B., WOLOSUIK, R. A. Photosynthesis: the carbon reactions. In: Taiz L, Zeiger E (ed) *Plant physiology*, 5th edn. Massachusetts, United States of America. Pp 199-242, 2010.

LUNN, J. E. et al. Purification, molecular cloning, and sequence analysis of sucrose-6P-phosphate phosphohydrolase from plants. **Proceedings of the National Academy of Sciences**, v. 97, n. 23, p. 12914-12919, 2000.

LUNN, J. E.; MACRAE, E. New complexities in the synthesis of sucrose. **Current opinion in plant biology**, v. 6, n. 3, p. 208-214, 2003.

LUNN, J. E.. Evolution of sucrose synthesis. **Plant physiology**, v. 128, n. 4, p. 1490-1500, 2002.

MARAFANTE, L. J. **Tecnologia da fabricação do álcool e do açúcar**. Icone, 1993.

OMETTO, A. R., (2000). Discussão sobre os fatores ambientais impactados pelo setor sucroalcooleiro e a certificação socioambiental. São Carlos, SP. Dissertação (Mestrado) –Escola de Engenharia de São Carlos, Universidade de São Paulo, 2000.

PHYTOZOME v11.0. Site: <https://phytozome.jgi.doe.gov/>.

RENNIE, E. A.; TURGEON, R. A comprehensive picture of phloem loading strategies. **Proceedings of the National Academy of Sciences**, v. 106, n. 33, p. 14162-14167, 2009.

ROBERTO, G. G. Fisiologia da maturação de cana-de-açúcar (*Saccharum* spp): Sinalização e controle do metabolismo de produção e armazenamento de sacarose. Campinas, SP. Tese de doutorado. Instituto agrônomo, 2015.

RUAN, Y-L. et al. Molecular regulation of seed and fruit set. **Trends in plant science**, v. 17, n. 11, p. 656-665, 2012.

RUAN, Y-L. Sucrose metabolism: gateway to diverse carbon use and sugar signaling. **Annual review of plant biology**, v. 65, p. 33-67, 2014.

SAMBROOK, J. et al. **Molecular cloning: a laboratory manual**. Cold spring harbor laboratory press, 1989.

SANTA BRIGIDA, A. B. et al. Sugarcane transcriptome analysis in response to infection caused by *Acidovorax avenae* subsp. *avenae*. **PLoS one**, v. 11, n. 12, p. e0166473, 2016.

SCHAKER, P. D. et al. RNAseq transcriptional profiling following whip development in sugarcane smut disease. **PLoS one**, v. 11, n. 9, p. e0162237, 2016.

SLEWINSKI, T. L.; BRAUN, D. M. Current perspectives on the regulation of whole-plant carbohydrate partitioning. **Plant Science**, v. 178, n. 4, p. 341-349, 2010.

SMITH, A. M.; COUPLAND, G.; DOLAN, L.; HARBERD, N.; JONES. J.; MARTIN, C.; SABLONSKI, R.; AMEY, A. *Plant Biology*. 1ª edição. 2010.

SOCICANA-Associação dos fornecedores de Cana de Guariba. Consultado em 2018. <http://socicana.com.br/>

SREENIVASAN, T. V.; AHLOOWALIA, B. S.; HEINZ, D. J. Sugarcane improvement through breeding. *Cytogenetics*. **Elsevier, Amsterdam**, p. 211-235, 1987.

SULPICE, R. et al. Starch as a major integrator in the regulation of plant growth. **Proceedings of the National Academy of Sciences**, v. 106, n. 25, p. 10348-10353, 2009.

TAIZ, L. et al. **Fisiologia e desenvolvimento vegetal**. Artmed Editora, 2017.

TOMKINS, J. P. et al. A bacterial artificial chromosome library for sugarcane. **Theoretical and Applied Genetics**, v. 99, n. 3-4, p. 419-424, 1999.

UNICA, Brazilian Sugarcane Industry Association. Industry. South Central 2015/2016 harvest should reach 590 million tons of processed cane, with priority on ethanol production, 2015.

VETTORE, A. L. et al. The libraries that made SUCEST. **Genetics and Molecular Biology**, v. 24, n. 1-4, p. 1-7, 2001.

WANG, J. et al. Carbon partitioning in sugarcane (*Saccharum* species). **Frontiers in plant science**, v. 4, p. 201, 2013.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973, 2007.

WHITAKER, D. P. Purification and properties of sucrose-6-phosphatase from *Pisum sativum* shoots. **Phytochemistry**, v. 23, n. 11, p. 2429-2430, 1984.