

ANA PAULA GOMES SOARES

**CARACTERIZAÇÃO *IN SILICO* DE GENES ENVOLVIDOS NA
BIOSSÍNTESE DE OLIGOSSACARÍDEOS DE RAFINOSE EM SOJA E
CONSTRUÇÃO DE CASSETES DE SILENCIAMENTO DO GENE DA
ESTAQUIOSE SINTASE, VIA INTERFERÊNCIA POR RNA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Agrícola, para obtenção do título de *Doctor Scientiae*.

**VIÇOSA
MINAS GERAIS – BRASIL
2011**

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

S676c
2011

Soares, Ana Paula Gomes, 1980-

Caracterização *in silico* de genes envolvidos na biossíntese de oligossacarídeos de rafinose em soja e construção de cassetes de silenciamento do gene da estaquiose sintase, via interferência por RNA / Ana Paula Gomes Soares. – Viçosa, MG, 2011.

xii, 123f. : il. (algumas col.) ; 29cm.

Texto em português e inglês.

Inclui anexos.

Orientador: Valéria Monteze Guimarães.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Enzimas. 2. Oligossacarídeos. 3. Soja. 4. Genes.
5. Genética molecular. 6. Carboidratos. 7. Bioquímica.
8. Biotecnologia. 9. Biologia molecular. I. Universidade
Federal de Viçosa. II. Título.

CDD 22. ed. 572.7

ANA PAULA GOMES SOARES

**CARACTERIZAÇÃO *IN SILICO* DE GENES ENVOLVIDOS NA
BIOSSÍNTESE DE OLIGOSSACARÍDEOS DE RAFINOSE EM SOJA E
CONSTRUÇÃO DE CASSETES DE SILENCIAMENTO DO GENE DA
ESTAQUIOSE SINTASE, VIA INTERFERÊNCIA POR RNA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 29 de abril de 2011.

Prof. Everaldo Gonçalves de Barros
(Co-Orientador)

Prof. Wagner Campos Otoni
(Co-Orientador)

Prof. Luiz Orlando de Oliveira

Prof^a. Cynthia Maria Borges Damasceno

Prof^a. Valéria Monteze Guimarães
(Orientadora)

*A meus pais João Batista e Graça,
pelo imenso amor, cuidado, incentivo e exemplo de vida.*

*Ao meu esposo, Renan,
pela dedicação, amor e alegria.*

*Aos meus irmãos, Leonardo e Leandro
pela amizade, carinho e apoio*

DEDICO

AGRADECIMENTOS

A Deus.

A minha orientadora, professora Valéria Monteze Guimarães, pela orientação, por todo apoio, confiança, ensino e amizade durante toda minha vida acadêmica.

Ao meu co-orientador, professor Everaldo Gonçalves de Barros, pelo apoio, pelas instruções e pela constante presença durante a execução deste trabalho.

Aos meus co-orientadores, professor Wagner Campos Otoni e professor Sebastião Tavares de Rezende, pelo apoio e pelas contribuições para realização deste trabalho.

Aos professores, Luiz Orlando de Oliveira e Cynthia Maria Borges Damasceno, pela participação na banca examinadora.

Aos professores do Departamento de Bioquímica e Biologia Molecular, em especial à professora Elizabeth Pacheco Batista Fontes e ao professor Luciano Gomes Fietto, pelo agradável convívio e contribuição em minha formação acadêmica.

Ao Eduardo Monteiro pelas informações e auxílio.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Bioquímica Agrícola, pela formação acadêmica.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

À Lídia, pela amizade e dedicação a este trabalho.

Aos amigos e colegas do Laboratório de Biologia Molecular e Fitogeografia (antigo Seq DNA), pelo companheirismo, amizade, e auxílio para obtenção dos resultados.

Às amigas Andrea, Daniela, Eliana, Mere, Josie e Silvana, pela presença constante em minha vida e aos amigos e irmãos da igreja pela atenção e pelas orações.

Aos funcionários da cozinha e almoxarifado, pela disposição constante em ajudar.

A todos aqueles que contribuíram, direta ou indiretamente, para o sucesso deste trabalho, o meu sincero agradecimento e reconhecimento.

BIOGRAFIA

ANA PAULA GOMES SOARES, filha de João Batista Soares e Maria das Graças Gomes Soares, nasceu no dia 22 de julho de 1980, na cidade de Ubá, Minas Gerais.

Em 1997, concluiu o ensino médio no Colégio Universitário – COLUNI – na Universidade Federal de Viçosa, na cidade de Viçosa – MG.

Em 2001, iniciou o curso de Bioquímica na Universidade Federal de Viçosa, concluindo-o em janeiro de 2005. Durante a graduação teve a oportunidade de realizar Iniciação Científica, desenvolvendo trabalhos de pesquisa por três anos no Laboratório de Biologia Molecular de Plantas, no Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO) – UFV.

Em 2005 ingressou no curso de mestrado do Programa de Pós-Graduação em Bioquímica Agrícola, do Departamento de Bioquímica e Biologia Molecular, da UFV. Durante esse período atuou na área de biologia molecular, no Laboratório de Biologia Molecular de Plantas, defendendo dissertação em fevereiro de 2007.

Ingressou, em março de 2007, no Programa de Pós-Graduação em Bioquímica Agrícola da UFV, em nível de doutorado, submetendo-se a defesa em 29 de abril de 2011.

CONTEÚDO

RESUMO	ix
ABSTRACT	xi
1. INTRODUÇÃO GERAL	1
2. REVISÃO BIBLIOGRÁFICA.....	4
2.1. Soja: aspectos gerais e importância econômica	4
2.1.1. Importância econômica e comercial da soja	4
2.1.2. Composição e aspectos nutricionais da soja	6
2.2. Biossíntese dos oligossacarídeos de rafinose.....	8
2.3. A enzima estaquiase sintase.....	12
2.4. Papel fisiológico dos oligossacarídeos de rafinose	16
2.5. Aspectos nutricionais dos oligossacarídeos de rafinose.....	18
2.6. Variedades de soja com reduzido conteúdo de ácido fítico e oligossacarídeos de rafinose.....	20
2.7. Silenciamento gênico pós-transcricional por RNA.....	23
2.7.1. Técnicas para indução do silenciamento gênico via RNAi.....	26
2.7.2. Principais avanços do uso do RNAi no melhoramento de plantas.....	27
3. REFERÊNCIAS BIBLIOGRÁFICAS.....	31
4. OBJETIVOS	49
4.1. Objetivos gerais.....	49
4.2. Objetivos específicos	49
CAPÍTULO 1	50
<i>IN SILICO</i> CHARACTERIZATION AND ANALYSIS OF GENE EXPRESSION INVOLVED IN THE BIOSYNTHESIS OF RAFFINOSE OLIGOSACCHARIDES IN SOYBEAN (<i>Glycine max</i> L. Merrill).....	50

ABSTRACT	50
1. INTRODUCTION.....	52
2. MATERIALS AND METHODS.....	54
3. RESULTS AND DISCUSSION	57
3.1. <i>In silico</i> identification and sequence analysis	57
3.2. Expression analysis	68
4. REFERENCES.....	75
SUPPLEMENTARY	81
CAPÍTULO 2.....	84
CONSTRUÇÃO DE CASSETES DE EXPRESSÃO PARA SILENCIAMENTO DO GENE DA ESTAQUIOSE SINTASE DE SOJA (<i>Glycine max</i> L. Merrill), VIA INTERFERÊNCIA POR RNA	84
RESUMO	84
1. INTRODUÇÃO	86
2. MATERIAL E MÉTODOS	88
2.1. Material Vegetal.....	88
2.2. Extração de RNA total	88
2.3. Construção de oligonucleotídeos iniciadores específicos e isolamento dos fragmentos diferentes do gene que codifica para estaquiose sintase para a construção do cassete de silenciamento	90
2.4. Amplificação dos fragmentos gênicos e clonagem no vetor pGEM T- <i>Easy</i>	91
2.5. Transformação de <i>Escherichia coli</i> por choque térmico e diagnóstico molecular dos transformantes	92
2.6. Clonagem dos fragmentos específicos em vetor pKANNIBAL	93
2.7. Nomenclatura dos clones contendo os fragmentos <i>sense</i> e <i>antisense</i> em vetor pKANNIBAL.....	94
2.8. Transferência das construções dos vetores pKANNIBAL para os vetores pCAMBIA 3301.....	94
2.9. Nomenclatura dos clones com as construções em vetor pCAMBIA 3301	95
2.10. Clonagem dos fragmentos específicos em vetor pBKN	96
2.11. Nomenclatura dos clones contendo as construções dirigidas pelo promotor do gene da subunidade α da β -concliginina em vetores pBKN.....	96
2.12. Transferência dos cassetes com o promotor do gene da subunidade α da β - congliginina do vetor pBKN para o vetor pCAMBIA 3301	96

2.13. Nomenclatura dos clones com as construções dirigida pelo promotor do gene da subunidade α da β -conglucininina em vetores pCAMBIA 3301	97
2.14. Clonagem do gene estaquiose sintase completo	97
3. RESULTADOS E DISCUSSÃO	99
3.1. Isolamento e clonagem do gene estaquiose sintase.....	99
3.2. Construção de cassetes de expressão para o silenciamento gênico da estaquiose sintase em soja.....	102
4. CONCLUSÕES	115
5. PERSPECTIVAS	116
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	117
ANEXO I	123

RESUMO

SOARES, Ana Paula Gomes, D. Sc., Universidade Federal de Viçosa, abril de 2011.
Caracterização *in silico* de genes envolvidos na biossíntese de oligossacarídeos de rafinose em soja e construção de cassetes de silenciamento do gene da estaquiase sintase, via interferência por RNA. Orientadora: Valéria Monteze Guimarães. Co-Orientadores: Everaldo Gonçalves de Barros, Wagner Campos Otoni e Sebastião Tavares de Rezende.

A soja (*Glycine max* L. Merrill) acumula oligossacarídeos de rafinose (ROs) em suas sementes, principalmente rafinose e estaquiase, os quais são considerados fatores antinutricionais. A redução destes carboidratos em sementes de soja poderia resultar na diminuição dos problemas gastrointestinais gerados a partir da ingestão de produtos derivados de soja e proporcionar um impacto positivo na aceitação dos mesmos. A biossíntese dos ROs é bem caracterizada e envolve a ação sequencial de um grupo de sintases. Entretanto, os genes e o padrão de expressão destas sintases ainda não foram completamente descritos em soja. O objetivo deste trabalho foi caracterizar e analisar a expressão *in silico* dos genes envolvidos na biossíntese de ROs e construir cassetes de expressão para silenciamento do gene da estaquiase sintase, via interferência por RNA, em soja. Análises computacionais mostram que o genoma da soja apresenta 10 locos que potencialmente codificam sintases de ROs. Destes, seis codificam a galactinol sintase, três a rafinose sintase e um a estaquiase sintase. Os perfis de expressão virtual dos putativos genes para galactinol sintase e rafinose sintase indicam que estes genes são expressos em sementes, folhas e raízes estressadas. Transcritos para estaquiase sintase são mais abundantes em folhas de plantas estressadas que em sementes. O genoma da soja também mostra sete locus que potencialmente codificam proteínas de embebição de sementes ou α -galactosidases alcalinas que são expressos principalmente em sementes. Para redução do conteúdo de ROs em sementes de soja, foram construídos cassetes para silenciamento, via interferência por RNA, do gene que codifica a

estaquiose sintase. Os fragmentos isolados contêm sequências correspondentes a duas regiões distintas do gene estaquiose sintase que apresentam alta e baixa identidade com genes de rafinose sintases de soja. Os fragmentos *sense* e *antisense* de cada região do gene foram clonados nos vetores pKANNIBAL e pBKN para a síntese simultânea de regiões repetidas e invertidas, resultando em RNA *hairpin* espaçados por um intron. Foram obtidas quatro construções nas quais duas são dirigidas pelo promotor 35SCaMV e duas pelo promotor da subunidade α da proteína de reserva β -conglucina. Estas construções foram transferidas para o vetor binário pCAMBIA 3301, para posterior transformação genética de soja.

ABSTRACT

SOARES, Ana Paula Gomes, D. Sc., Universidade Federal de Viçosa, April, 2011. ***In silico* characterization of a gene involved in the biosynthesis of raffinose oligosaccharides in soybean and construction of cassettes for silencing of stachyose synthase gene, via RNA interference.** Adviser: Valéria Monteze Guimarães. Co-Advisers: Everaldo Gonçalves de Barros, Wagner Campos Otoni and Sebastião Tavares de Rezende

Soybean (*Glycine max* L. Merrill) seed accumulates raffinose oligosaccharides (ROs), mainly raffinose and stachyose, which are considered antinutritional factors. The reduction of these carbohydrates in soybean seeds could result in decreased gastrointestinal problems generated from soy products and provide a positive impact on acceptance of these products. The biosynthesis of ROs is well characterized and involves the sequential action of a group of synthases. However, genes and the pattern of expression of these synthases have not been fully described in soybean. The aim of this study was to characterize and analyze *in silico* the expression of genes involved in biosynthesis of ROs and construction expression cassettes for gene silencing of stachyose synthase via RNA interference in soybean. Computer analysis showed that the soybean genome has 10 *loci* which potentially encode for synthases ROs. Of these, six encode galactinol synthase, three encode raffinose synthase and one encodes stachyose synthase. The expression profiles of candidate genes for galactinol synthase and raffinose synthase suggest that these genes are expressed in seeds, leaves and roots stressed. Stachyose synthase transcripts showed to be more abundant in leaves of stressed plants than in seeds. The genome of soybean also showed seven *loci* which potentially encode for seed inhibition proteins or alkaline α -galactosidases expressed mainly in seeds. To reduce the ROs content in soybean seeds were constructed expression cassettes for silencing of the gene encoding stachyose synthase, via RNA

interference. The fragments isolated contain sequences corresponding to two distinct regions of stachyose synthase gene that show high and low identity with the raffinose synthase gene from soybean. The sense and antisense fragments of each gene region were cloned into the vectors pKANNIBAL and pBKN for simultaneous synthesis of inverted repeat regions, resulting in RNA hairpin spaced by an intron. Four constructions were obtained in which two are directed by the promoter 35SCaMV and two are directed by the promoter of the subunit of storage protein β -conglycinin. These constructs were transferred for binary vector pCAMBIA 3301 for further soybean genetic transformation.

1. INTRODUÇÃO GERAL

A soja (*Glycine max* L. Merrill) é uma leguminosa cultivada mundialmente, correspondendo à quarta cultura em produção de grãos no mundo. O Brasil é o segundo maior produtor mundial dessa leguminosa, sendo a produção na safra 2010/2011 estimada em aproximadamente 70,1 milhões de toneladas, superior em 2,1% (1,41 milhão de toneladas) a safra de 2009/2010 (CONAB, 2011a). Esta oleaginosa é a principal cultura do agronegócio brasileiro, sendo a receita das exportações do complexo da soja (em grão, farelo e óleo), nesse período, de US\$ 17,1 bilhões, o que corresponde a 23% do total de exportação do agronegócio brasileiro e a 8,5% do total nacional de exportação (CONAB, 2011b).

A importância econômica da soja se deve, principalmente, ao seu elevado valor nutricional e as suas propriedades funcionais. As sementes de soja contêm alto conteúdo de lipídeos (15-25%), de proteínas (30-45%) e de carboidratos (20-35%) (Deleprane et al., 2009). A maior parte destes carboidratos (aproximadamente 99%) corresponde à sacarose e oligossacarídeos de rafinose (ROs) (Kumar et al., 2010).

Em sementes maduras de soja, a presença de ROs, em altas concentrações, constitui um fator antinutricional para humanos e animais monogástricos, restringindo o consumo desta leguminosa e de seus derivados na alimentação. O trato intestinal humano e de animais monogástricos não possui a enzima α -galactosidase (EC 3.2.1.22, α -D- galactosilgalactohidrolase), necessária para hidrolisar as ligações α -(1 \rightarrow 6) dos resíduos de galatosil presentes nestes açúcares. Conseqüentemente eles não são digeridos causando disfunções orgânicas, tais como diarreia, náuseas e flatulência (Liener, 1994).

Em plantas, a via metabólica de síntese de ROs já está bem estabelecida. A síntese destes oligossacarídeos depende da ação sequencial de uma série de

galactosiltransferases e galactosilhidrolases. No primeiro passo, UDP-galactose e mio-inositol formam galactinol mais UDP, pela atividade da enzima galactinol sintase (Gols). No segundo passo, a enzima rafinose sintase (RFS) transfere um resíduo de galactose do galactinol para a sacarose, formando rafinose e mio-inositol. No terceiro passo, a enzima estaquiase sintase (STS) transfere outro resíduo de galactose do galactinol para a rafinose, por meio de uma ligação α -(1-6), liberando os produtos estaquiase e mio-inositol. Posteriormente são sintetizadas verbascose e ajucose, por ação da verbascose e ajucose sintase, respectivamente (Peterbauer et al., 2002a).

Os ROs são comumente acumulados durante o desenvolvimento e maturação das sementes de leguminosas (Peterbauer e Richter, 1998), desempenhando importantes papéis fisiológicos nas mesmas. Embora não sejam vitais (Dierking e Bilyeu, 2009a), estes oligossacarídeos servem como fonte primária de energia durante o processo de germinação das sementes (Lahuta et al., 2000); exercem importante papel no processo de tolerância contra a dessecação durante a maturação da semente (Wu et al., 2009); foram positivamente relacionados à longevidade da semente durante o período de armazenamento (Haagenson, 2008); e estão associados com respostas fisiológicas e bioquímicas a condições de estresse em plantas (Taji et al., 2002). Além disso, eles são descritos como importantes carboidratos de transporte no floema em algumas espécies vegetais (Sprenger e Keller, 2000; Voitsekhovskaja et al., 2009).

O potencial da rafinose e da estaquiase presentes na soja em gerar problemas gastro-intestinais e, conseqüentemente, limitar a utilização destes grãos na alimentação, estimulou o desenvolvimento de métodos que proporcionam a redução do conteúdo de ROs em produtos derivados de soja, destacando-se técnicas como a hidrólise enzimática (Guimarães et al., 2001), hidratação dos grãos (Sanni et al., 1997), cozimento (Pugalenthi et al., 2006), fermentação (Rehms e Braz, 1995) e extração dos ROs com água (Ku et al., 1976) e com etanol (Martínez-Villaluenga et al., 2006).

Em programas de melhoramento, variedades de soja com reduzido teor de ROs foram obtidas, principalmente, através de variedades de soja que apresentam modificações nos genes que codificam a enzima mio-inositol-1-fosfato sintase (MIPS) e a enzima RFS (Sebastian et al., 2000; Hitz et al., 2002, 2004; Nunes et al., 2006; Frank et al., 2009; Yuan et al., 2009; Dierking e Bilyeu, 2009b). A enzima MIPS atua no início da via de biossíntese dos ROs, e alterações na atividade da mesma ou no padrão de expressão do gene resultam em mudanças da quantidade de todos os intermediários da via, enquanto que a alteração da atividade da RFS resultou apenas na mudança do

conteúdo de rafinose e estaquiose nas sementes de soja, que embora reduzidos não foram eliminados (Dierking e Bilyeu, 2009b).

Considerando que rafinose e estaquiose são não essenciais para a germinação das sementes de soja (Dierking e Bilyeu, 2008), outra estratégia para se obter soja com reduzido teor de ROs, é o silenciamento dos genes que codificam para rafinose e/ou estaquiose sintase. O silenciamento destes genes pode reduzir a síntese destes oligossacarídeos e contribuir para obtenção de variedades de soja com menor conteúdo de ROs.

Uma técnica que vem sendo amplamente empregada para o silenciamento de genes específicos é a interferência por RNA ou RNA interference (RNAi). Este método de silenciamento gênico pós-transcricional é altamente eficiente, principalmente quando a construção resulta na formação de um RNA *hairpin* espaçado por um íntron (hpRNA) (Wesley et al., 2001).

Para a seleção dos genes a serem silenciados é necessário determinar, inicialmente, os genes constituintes da via de biossíntese de ROs e analisar o padrão de expressão gênica dos mesmos. A partir do sequenciamento do genoma da soja (<http://www.phytozome.net/>) e sua publicação recente (Schmutz et al., 2010), os processos de identificação, clonagem e análise de expressão de genes podem ser rapidamente realizados e utilizados como ferramentas para o desenvolvimento de variedades de soja com constituintes nutricionais alterados, tais como proteínas, óleo e carboidratos.

Este trabalho está organizado em dois capítulos. O primeiro capítulo está em formato de artigo, e será submetido à revista *Plant Cell Reports*, e apresenta a identificação, caracterização e a análise da expressão *in silico* de genes envolvidos na biossíntese de ROs em soja. O segundo capítulo, escrito no formato livre, descreve a construção de quatro cassetes de expressão visando ao silenciamento do gene da estaquiose sintase em soja, por RNA de interferência, uma vez que esta estratégia pode ser utilizada como ferramenta molecular para a obtenção de variedades de soja transgênicas apresentando baixos teores de ROs.

2. REVISÃO BIBLIOGRÁFICA

2.1. Soja: aspectos gerais e importância econômica

2.1.1. Importância econômica e comercial da soja

A soja corresponde à quarta cultura em produção de grãos e é a oleaginosa mais cultivada no mundo, sendo, por isso, freqüentemente denominada de “cultura milagre” (L’Hocine e Boye, 2007). De acordo com a Organização das Nações Unidas para Agricultura e Alimentação (FAO), em 1990 a produção mundial de soja em grão foi cerca de 108 milhões de toneladas; em 2009 atingiu o patamar de 260 milhões; e a safra de 2010/2011 está projetada para aproximadamente 256 milhões de toneladas. Embora as estatísticas apontem para uma redução de 1,6% da produção da safra de 2010/2011 em relação a safra anterior (2009/2010), também demonstram que houve aumento de 137% na produção mundial de soja, ao longo de duas décadas.

Este aumento da produção é acompanhado por um aumento no consumo. Dados do Departamento de Agricultura dos Estados Unidos (USDA - *United States Department of Agriculture*) indicam que apenas nos últimos três anos o consumo mundial de soja aumentará 15%, de 221,1 milhões de toneladas em 2009 para 255,5 milhões de toneladas em 2011 (previsão).

Dentre os países produtores de soja se destacam os Estados Unidos, o Brasil, a Argentina e a China, os quais foram responsáveis, em 2009/2010, por 87% da produção mundial de soja, enfatizando que os Estados Unidos obtiveram a maior produção no contexto mundial (35%). O Brasil, segundo colocado (27%), vem aumentando gradativamente a produção desta cultura, seguido pela Argentina (21%) e pela China (6%) (USDA, 2011).

A soja é a cultura mais plantada no país, apresentando ampliação da área cultivada de 10,6 milhões, em 1995, para 23,46 milhões de hectares, em 2009/2010 (CONAB, 2011a). Este aumento de 121% na área cultivada é acompanhado por um aumento de produção de 196% nos últimos 12 anos, passando de 23,2 milhões, em 1995, para 68,7 milhões de toneladas na safra 2009/2010 (CONAB, 2011a).

O valor da produção brasileira em 2010/2011 está estimado para 70,1 milhões de toneladas, o que corresponde a um aumento de 2,1% na produção deste grão em relação a safra de 2009/2010. Segundo a Companhia Nacional de Abastecimento (CONAB-2011a), esse aumento na produção se deve a expansão da área plantada, que ficou 2,8% maior, ocupando, 24,1 milhões de hectares no Brasil em 2010/2011, e a regularidade das chuvas pelo país.

O Mato Grosso é o maior produtor brasileiro de soja com 27,3% do total nacional na safra 2009/2010, com uma produção de 18,76 milhões de toneladas. O Paraná vem em segundo lugar respondendo por 20,5%, do total da safra, com a produção de 14,07 milhões de toneladas. Minas Gerais ocupa a sétima posição respondendo por 4,2% do total da safra brasileira em 2009/2010 (CONAB, 2011a).

Embora o grão de soja inteiro seja utilizado para alimentação em alguns países orientais, o principal uso de soja na alimentação é baseado no tratamento e fracionamento desta em farinhas e óleos, especialmente nos países ocidentais. Desde a década de 1960 novos produtos derivados da soja têm sido desenvolvidos. Os derivados protéicos da soja, para a alimentação, se enquadram em três produtos básicos: farinhas, com cerca de 50% de proteínas; concentrados protéicos de soja, com 70% de proteína e isolados protéicos de soja com 90-97% de proteína. Estes produtos podem sofrer processamento originando produtos de soja texturizados que são utilizados em panificadoras como clareadores de farinhas e agentes para melhoria das características de panificação e conservação (Moreira, 1999). O extrato hidrossolúvel de soja é utilizado em substituição ao leite de origem animal e é indicado principalmente para pessoas com deficiência genética, incapazes de metabolizar lactose (De Lumen, 1992). Para alimentação animal, a forma mais comum de uso dos produtos protéicos de soja é como farelo. O farelo de soja é obtido como produto final após extração do óleo de soja e possui de 10-12% de umidade e, em média, 48% de proteínas (Goldflus, 2001).

A soja é a principal oleaginosa produzida no Brasil, sendo assim, ela responde por 95% do mercado de óleo vegetal brasileiro para alimentação humana, além de ser o principal componente para produção de biodiesel no Brasil na atualidade (Osaki e

Batalha, 2008). A soja também é utilizada em diversos setores industriais, incluindo aplicações cosméticas e farmacêuticas (L'Hocine e Boye, 2007) e tem sido estudada como fonte de substâncias denominadas fitoquímicos, entre os quais estão os flavonóides. Neste sentido, foi observada uma relação entre o consumo de soja e a redução dos riscos de doenças crônicas não-infecciosas como as doenças cardiovasculares, alguns tipos de câncer e osteoporose (Messina e Messina, 1991; Lichtenstein, 1998; Morais e Silva, 2000; Tavares et al., 2000; Esteves e Monteiro 2001).

2.1.2. Composição e aspectos nutricionais da soja

A importância econômica da soja se deve, principalmente, ao seu elevado valor nutricional e as suas propriedades funcionais. A soja é excelente fonte de proteínas, óleos poliinsaturados, fibras, vitaminas, minerais e outros nutrientes, tanto para a alimentação humana quanto para alimentação animal (Krishna, 2005; Zarkadas, 2007). A composição do grão de soja pode variar dependendo da variedade e das condições de crescimento (Kumar et al., 2010). Os grãos das variedades de soja cultivadas apresentam uma composição média de 17% de óleo, 35% de proteínas e 31% de carboidratos em relação à massa seca (L'Hocine e Boye, 2007).

O conteúdo de proteína do grão de soja é maior do que o conteúdo das fontes protéicas normalmente consumidas na alimentação humana, tais como peixes, queijos, ovos e leite. A soja contém em suas proteínas quantidade suficiente de quase todos os aminoácidos essenciais, com exceção de metionina (Krishnan, 2005).

Os lipídeos da soja (óleo de soja cru) são constituídos principalmente de triacilgliceróis (97,9%). Os triacilgliceróis podem variar em ácidos graxos constituintes e sua composição depende da variedade vegetal (Neff et al., 1992). O conteúdo médio dos ácidos graxos constituintes do óleo de soja das variedades comerciais é de 4% de ácido esteárico, 7% de ácido linolênico, 11% de ácido palmítico, 24% de ácido oléico, 54% de ácido linoléico. Este elevado teor de ácidos graxos polinsaturados (61%) presente no óleo de soja tem sido relacionado com a redução da estabilidade oxidativa e, conseqüentemente, da qualidade do óleo, produzindo compostos de reduzido peso molecular, voláteis, responsáveis pelo *off-flavor*, rancidez e desempenho reduzido do óleo (Yadav, 1996).

Os carboidratos são o segundo grupo mais abundante presente na semente da soja (31%), indicando um importante valor nutricional e econômico para este grupo na indústria alimentícia. Estes carboidratos são representados por açúcares solúveis, tais como sacarose, estaquiose e rafinose, e por fibras insolúveis, que incluem principalmente hemicelulose, celulose e pectina. Segundo Hymowitz e colaboradores (1972a) a sacarose e oligossacarídeos de rafinose (ROs), principalmente rafinose e estaquiose, correspondem a aproximadamente 99% dos carboidratos solúveis presentes na semente de soja madura.

O conteúdo de sacarose e ROs pode variar de acordo com o cultivar e com as condições de crescimento. Kumar e colaboradores (2010), analisando o conteúdo de sacarose e RO em 148 genótipos de soja, observaram que há maior variação do teor de sacarose (4,8 vezes) em relação ao teor de ROs (2,63 vezes) nos genótipos analisados. As análises, realizadas por cromatografia líquida de alto desempenho (HPLC – *High Performance Liquid Chromatography*), revelaram também variação de 0,64-2,53 mmol.100g⁻¹ e 2,09-7,1 mmol.100g⁻¹ no conteúdo de rafinose e estaquiose, respectivamente, entre todas as sementes das variedades de soja analisadas. Trugo e colaboradores (1995), analisando 20 variedades de soja, verificaram variação de 3,7-5,6% da massa seca das sementes, no conteúdo de sacarose. Nestas sementes, o teor total de ROs variou de 3,9 a 5,3% da massa seca, sendo que o conteúdo de estaquiose foi maior em todas as sementes analisadas (4,6%). Egounlety e colaboradores (2003), analisando o conteúdo de sacarose e ROs em sementes de soja variedade TGX536-02D, observaram que o teor de sacarose (4,91% da massa seca das sementes) foi maior que o de ROs (4,63%). Nesta variedade, assim como nas demais, o conteúdo de estaquiose (3,41% da massa seca das sementes) foi maior do que o de rafinose (1,22%).

De acordo com Fialho (2007), em sementes de soja, variedade CAC-1, a sacarose foi o açúcar mais abundante (4,7% da massa seca das sementes). A rafinose está presente nas sementes em concentrações muito baixas, atingindo máxima concentração na soja madura de 0,53%. A estaquiose, encontrada apenas na soja madura, apresentou uma concentração de 4,1%. O acúmulo de rafinose e, especialmente, estaquiose nos últimos estágios de desenvolvimento da semente também foi observado na variedade Chippewa 64 (Obendorf et al., 1998).

2.2. Biossíntese dos oligossacarídeos de rafinose

A soja acumula ROs durante a maturação das sementes (Pridham e Hassid, 1965; Obendorf et al., 1998). A síntese destes oligossacarídeos requer sacarose, na qual são adicionados unidades de galactose por ligação α -1,6 formando a rafinose, a partir da qual novas unidades de galactose são adicionadas, por ligação α -1,6 ao terminal galactose, para formar os tetra-, penta- e hexassacarídeos: estaquiose, verbascose e ajucose, respectivamente (Figura 1). Portanto, a biossíntese de ROs progride pela transferência de grupos galactosil originados de um doador galactosil apropriado (UDP-D-galactose, galactinol [*O*- α -D-galactopiranosil-(1 \rightarrow 1)-L-*mio*-inositol] ou dos próprios ROs, para uma molécula aceptora, a qual pode ser sacarose, rafinose, estaquiose ou verbascose (Peterbauer e Richter, 2001).

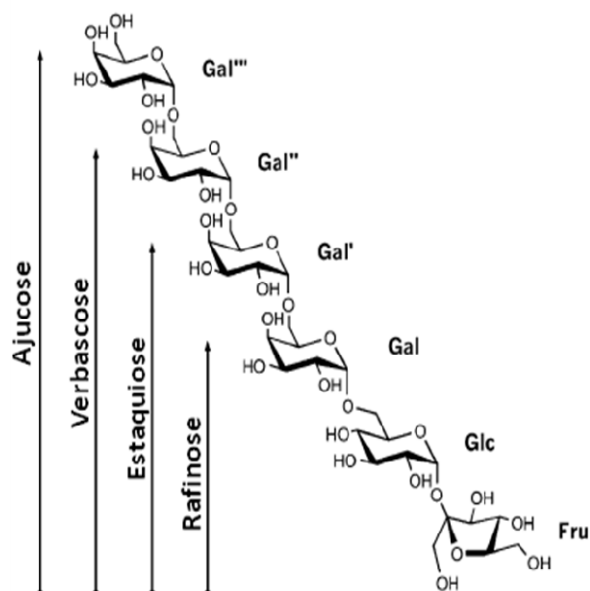


Figura 1 Estrutura dos açúcares envolvidos na síntese de ROs (Adaptado de Kotiguda et al., 2006).

O passo inicial para biossíntese de ROs é a formação de galactinol (Figura 2). O mio-inositol é fundamental para síntese de galactinol e é sintetizado pelas enzimas mio-inositol-1-fosfato sintase (MIPS; EC 5.5.1.4) (Hitz et al., 2002) e mio-inositol monofosfatase (IMP1; EC 3.1.3.25) (Styer et al., 2004). A UDP-D-galactose não é utilizada diretamente para a síntese de rafinose, mas representa o primeiro ponto de entrada de resíduos galactosil para a síntese de ROs. Neste sentido, o resíduo galactosil da UDP-D-galactose é transferido para o mio-inositol, produzindo galactinol, um doador de grupos galactosil específico para ROs. A síntese de rafinose [*O*- α -D-galactopiranosil-

(1→6)-α-D-glucopiranosil-(1→2)-β-D-fructofuranoside] corresponde a próxima etapa da via, que pode ser representada como se segue (Dey, 1985; Peterbauer et al., 2002a):

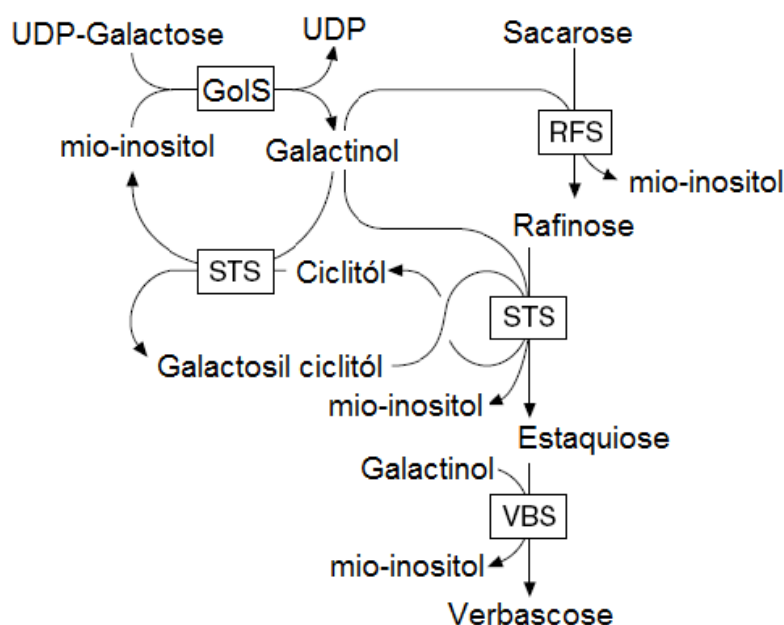
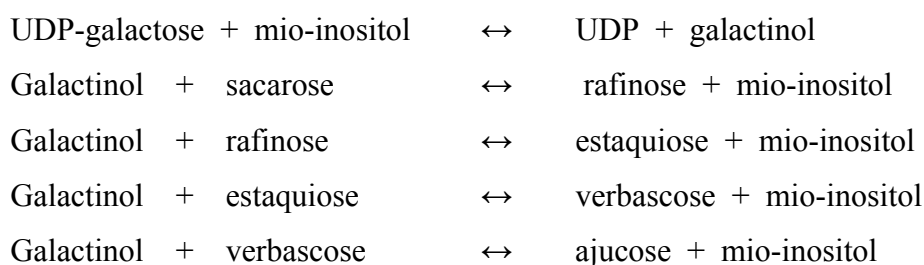


Figura 2 Representação esquemática da via de biossíntese dos ROs. GolS, galactinol sintase; RFS, rafinose sintase; STS, estaquiose sintase; VBS, verbascose sintase. Todas as reações são reversíveis (Adaptado de Peterbauer e Richter, 2001).

A síntese de galactinol é catalisada pela enzima galactinol sintase (GolS; UDP-galactose:mio-inositol 3-α-D-galactosiltransferase; EC 2.4.1.123) que produz galactinol a partir de UDP-galactose e mio-inositol (Saravitz et al., 1987; Obendorf et al., 2004). A enzima GolS é frequentemente considerada como ponto de regulação para biossíntese dos ROs uma vez que a principal função conhecida do galactinol é como substrato para a formação de ROs.

Segundo Castillo e colaboradores (1990) esta enzima parece auxiliar na regulação dos níveis dos oligossacarídeos de reserva nos órgãos vegetais de leguminosas, pois estudos demonstram que a atividade da galactinol sintase está

positivamente correlacionada com níveis de ROs durante o desenvolvimento das sementes (Haagenson et al., 2008).

A GolS foi purificada a partir de folhas de abobrinha de moita (*Cucurbita pepo*) (Webb, 1982; Smith et al., 1991) e de cotilédones de feijão (*Phaseolus vulgaris*) (Liu et al., 1995). Baseado nestes estudos, a GolS parece ser uma enzima monomérica com massa molecular entre 38-43 kDa, que requer Mn^{+2} para sua atividade. Evidências mostram ser uma enzima extra-vacuolar, provavelmente localizada no citossol (Bachamann e Keller, 1995). Em soja esta enzima ainda não foi purificada, entretanto Ribeiro e colaboradores (2000) descreveram a purificação parcial, a atividade enzimática e algumas propriedades cinéticas de uma galactinol sintase de sementes de soja.

Pelo menos um gene que codifica para GolS já foi identificado em diferentes espécies de plantas, tais como abobrinha (Kerr et al., 1993), *Zea mays* (Zhao et al., 2004), *Arabidopsis thaliana* (Lin et al., 1999), *Ajuga reptans* (Sprenger e Keller, 2000), *Solanum lycopersicum* (Downie et al., 2003), *Coptis japonica* (Takanashi et al., 2008) e *Populus trichocarpa* (Philippe et al., 2010). Em soja, foi descrito, até o momento, apenas o isolamento de um cDNA de galactinol sintase de soja (*GmGolS*) e sua expressão heteróloga em *Escherichia coli* (Obendorf et al., 2004).

A expressão dos genes *GolS* em respostas a diferentes tipos de estresses abióticos tem sugerido um importante papel para os ROs na adaptação aos mesmos. Segundo Taji e colaboradores (2002), em *A. thaliana*, os genes *GolS* são induzidos por baixas temperaturas, altas concentrações salinas e por deficiência hídrica. Estes autores afirmam que galactinol e rafinose estão envolvidos nos processos de tolerância a diferentes tipos de estresse abiótico. Em plantas de *A. thaliana*, a expressão de dois genes que codificam para galactinol sintase (*AtGolS1* e *AtGolS2*) foi induzida em resposta a estresse hídrico e salino, mas não por baixas temperaturas. Entretanto, outro gene da mesma família (*AtGolS3*) foi induzido apenas durante estresse causado por baixas temperaturas. Neste trabalho também foi demonstrado que a superexpressão de *AtGolS2* aumentou a tolerância à seca em plantas transgênicas de *Arabidopsis*, pois causou aumento no conteúdo de galactinol e rafinose endógenos e promoveu a redução da transpiração.

Em café (*Coffea arabica*), três isoformas de GolS (GolS1, GolS2 e GolS3) tiveram sua expressão alterada em resposta a diferentes estresses abióticos. GolS1 foi a isoforma mais expressa durante todos os tratamentos de estresses (hídrico, salino e por

calor). GolS2 foi detectada exclusivamente em condições de deficiência hídrica e estresse salino severos, enquanto GolS3 foi identificada tanto sob condições moderadas quanto sob condições severas de deficiência hídrica (Santos et al., 2011).

O bloqueio da expressão do gene *Gols* promoveria a redução do fluxo de carbono na via, inibindo parcial ou completamente a síntese dos ROs nas plantas. A redução da expressão ou da atividade desta enzima poderia diminuir os problemas nutricionais associados com os ROs, mas poderia resultar também em plantas e/ou sementes mais susceptíveis ao estresse ambiental.

Subseqüentemente à produção de galactinol, sintases específicas catalisam a síntese de cada membro da série. A rafinose é formada pela ação da enzima rafinose sintase (RFS; α -D-galactosyl-(1 \rightarrow 3)-1D-*mio*-inositol:sucrose 6- α -D-galactosyltransferase; EC 2.4.1.82) (Lehle e Tanner, 1973) e a estaquiase (β -D-fructofuranosil-O- α -D-galactosilpiranosil-(1 \rightarrow 6)- α -Dgalactopiranosideo-(1 \rightarrow 6)- α -D-glucopiranosídeo) é sintetizada pela enzima estaquiase sintase (STS; α -D-galactosil-(1 \rightarrow 3)-1-D-*mio*-inositol:rafinose galatosiltransferase; EC 2.4.1.67) (Peterbauer e Richter, 1998).

A RFS é uma enzima que catalisa reversivelmente a transferência de um resíduo galactosil de uma molécula de sacarose para o galactinol produzindo rafinose e mio-inositol (Figura 2). O exato papel desta enzima durante o desenvolvimento e germinação das sementes não foi completamente esclarecido, mas a produção desta enzima tem demonstrado ser importante para adaptação de plantas a estresses abióticos (Nishizawa et al., 2008; Lahuta e Górecki, 2010).

O conhecimento sobre as características bioquímicas da RFS ainda é limitado. Embora a RFS tenha sido purificada cerca de 400 vezes a partir de sementes de fava (*Vicia faba*) (Lehle e Tanner, 1973), esta enzima foi caracterizada apenas a partir do extrato bruto de folhas de *A. reptans* (Bachamann et al., 1994) e suas propriedades cinéticas foram relatadas a partir da enzima parcialmente pura originada de sementes de ervilha (*Pisum sativum*) (Peterbauer et al., 2002a,b). A RFS tem sido descrita com uma enzima de cerca de 90 kDa, que exibe um pH ótimo entre 6,5 e 7,0 e para qual os valores de K_m aparente para galactinol e sacarose estão na faixa de milimolar (Peterbauer et al., 2002b).

Recentemente foi descrita uma RFS de arroz (*Oryza sativa*) que, depois de clonada e expressa em *E. coli*, demonstrou ser ativa em pH 7,0 e a 45°C. Esta enzima foi específica para galactinol e p-nitrofenol- α -D-galactosídeos como doadores de grupos

galactosil, e sacarose, lactose, 4- β galactose, trealose lacto-*N*-biose foram reconhecidos como aceptores de grupos galactosil (Li et al., 2007).

Genes *RFS* foram identificados em algumas espécies vegetais, tais como *A. thaliana* (Kotani et al., 1997), *P. sativum* (Peterbauer et al., 2002a), *O. sativa* (Li et al., 2007) e *G. max* (Dierking e Bilyeu, 2008); entretanto a confirmação bioquímica e a caracterização molecular foi realizada completamente apenas para RFS de ervilha (Peterbauer et al., 2001; Peterbauer et al., 2002a)

A síntese de verbascose é catalisada por uma enzima denominada verbascose sintase. Entretanto, ainda não foi descrita a purificação e/ou caracterização de nenhuma enzima que execute apenas atividade de verbascose sintase. Peterbauer e colaboradores (2002b) purificaram, a partir de sementes de ervilha, uma estaquiose sintase multifuncional (EC 2.4.1.67) que além da síntese de estaquiose, catalisa a síntese de verbascose pela transferência de resíduos galactosil do galactinol para estaquiose, bem como pela transferência de resíduos de galactose terminais de uma molécula de estaquiose para outra, em pH 7,0.

A ajucose é um hexassacarídeo raramente encontrado em leguminosas. Kotiguda e colaboradores (2006) identificaram, por cromatografia de camada delgada (TCL - *Thin Layer Chromatography*) e cromatografia de papel, a presença de ajucose em sementes de feijão-da-índia (*Vigna mungo* L.). Este ROs foi isolado por cromatografia de sílica gel, e a análise de sua estrutura foi realizada através de hidrólise ácida, espectrometria de massa por bombardeamento com átomos rápidos (FABMS - *Fast atom bombardment mass spectrometry*) e por técnicas de ressonância magnética nuclear (NMR - *Nuclear magnetic resonance*). Entretanto, ainda não foi caracterizada e/ou purificada nenhuma enzima que apresente atividade de ajucose sintase em plantas.

2.3. A enzima estaquiose sintase

A enzima estaquiose sintase (STS; EC 2.4.1.67) é uma galactinol-rafinose galactosiltransferase que transfere um resíduo galactosil do galactinol para a rafinose (Figura 2), produzindo estaquiose e mio-inositol. Evidências cinéticas indicam que STS exibe um mecanismo de substituição enzimática, no qual todas as reações parciais são prontamente reversíveis. Em um primeiro passo, galactinol e STS formam um complexo enzima-galactose e mio-inositol é liberado como produto. Em um segundo passo, o

complexo enzima-galactose reage com a rafinose para produzir estaquiase (Richter et al., 2000; Peterbauer e Richter, 2001).

Esta enzima foi originalmente descrita em sementes de feijões (*P. vulgaris*) (Tanner e Kandler, 1968), mas foi purificada pela primeira vez em folhas maduras de abobrinha (Gaudreault e Webb, 1981). A STS de abobrinha foi purificada usando precipitação com sulfato de amônio, cromatografia de gel filtração e troca iônica. A enzima purificada apresentou atividade ótima em pH 6,9, foi particularmente inibida pelos íons Co^{+2} , Hg^{+2} , Mn^{+2} , Ni^{+2} e apresentou alta especificidade para o doador galactinol (K_m 7,7 mM) e para o acceptor rafinose (K_m 4,6 mM), sendo incapaz de sintetizar efetivamente qualquer outro membro da série dos ROs (Gaudreault e Webb, 1981). A partir deste trabalho, diversos estudos de purificação e caracterização da enzima STS têm sido realizados.

A enzima STS foi isolada, purificada e caracterizada a partir de folhas maduras de melão (*Cucumis melo*) (Holthaus e Schmitz, 1991a). Esta enzima é uma proteína com valor de pI em pH 5,1, e, apresentou duas subunidades de massa molecular de 45 e 50 kDa. A enzima apresentou atividade ótima em pH 6,8 na temperatura de 32°C. O Valor de K_m para a rafinose foi de 3,3 mM e para o galactinol foi de 7,7 mM. O mio-inositol foi um forte inibidor competitivo com valor de k_i de 1,8 mM. Holthaus e Schitz (1991b) também descreveram a distribuição e a imunolocalização desta enzima em plantas de melão, e concluíram que a síntese de estaquiase parece ocorrer nas veias foliares menores. Este resultado foi reafirmado em trabalhos posteriores com folhas de *C. pepo*, de oliva (*Olea europea* L.) e de *Alonsoa meridionalis* (Beebe e Turgeon, 1992; Flora e Madore, 1993; Voitsekhovskaja et al., 2009).

Pesquisas têm estabelecido que as plantas que possuem grande quantidade de conectores simplásticos entre o mesófilo e o floema, tal como cucurbitaceae e lamiaceae, sempre transportam, além de sacarose, ROs, sendo que a síntese destes ocorre nas células companheiras de nervuras foliares de ordens inferiores (Holthaus e Schitz 1991b; Beebe e Turgeon, 1992; Flora e Madore, 1993; McCaskill e Turgeon, 2007; Voitsekhovskaja et al., 2009). Entretanto, segundo Dey (1985), na maioria das espécies vegetais a síntese de estaquiase ocorre principalmente nas sementes.

Em folhas de *A. reptans* L., a STS apresentou atividade máxima em pH entre 6,5 e 7,0 e temperatura de 30 °C. O valor de K_m para o galactinol foi 2,4 mM na presença de 10 mM de rafinose e 4,8 mM para rafinose na presença de 5 mM de galactinol (Bachmann et al., 1994). Em sementes do feijão adzuki (*Vigna angularis*), a STS, sob

condições desnaturantes, revelou ser um único polipeptídeo de 90 kDa e foi separada em algumas formas ativas por focalização isoelétrica e cromatofocalização, com valores de pI entre pH 4,7 e 5,0. A STS purificada catalisou a transferência do grupo galactosil do galactinol para a rafinose formando mio-inositol e estaquiose como produto. O valor de K_m para rafinose na presença de galactinol foi 38,6 mM. A enzima, além de sintetizar estaquiose a partir de galactinol e rafinose, foi capaz de sintetizar galactopinitol a partir de D-pinitol e galactinol (Figura 3), como também foi capaz de realizar a reação inversa sintetizando estaquiose e D-pinitol a partir de rafinose e galactopinitol. Estes resultados demonstram que a enzima STS atuou sobre múltiplos substratos, e ocupa um papel central em reações de transferência de galactose (Peterbauer e Richter, 1998; Obendorf et al., 2004).

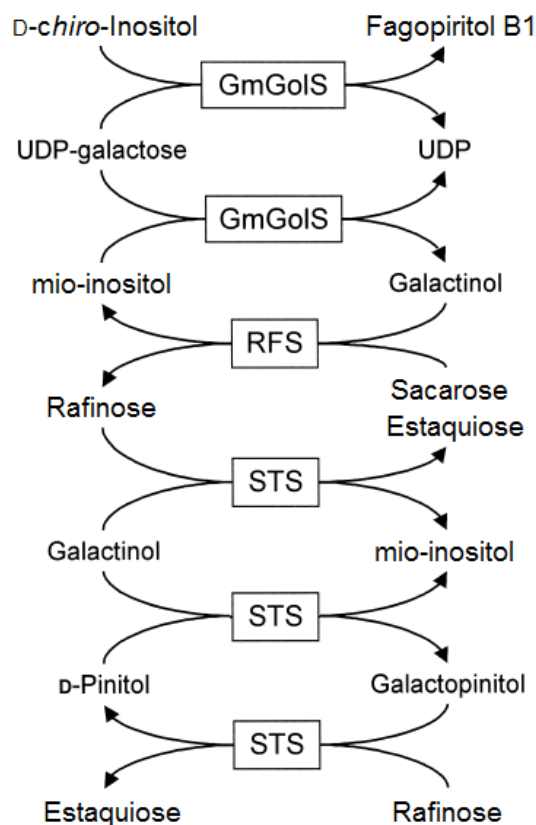


Figura 3 Representação esquemática da via de biossíntese de galactinol, rafinose, estaquiose, galactononitol e galactopinitol. GmGolS, *G. max* galactinol sintase; RFS, rafinose sintase; STS, estaquiose sintase (Adaptado de Obendorf et al., 2004).

Também em sementes maduras de lentilha (*Lens culinaris*), a STS foi purificada 313 vezes e apresentou ser uma proteína monomérica com massa molecular de 88,6 kDa, por SDS-PAGE, com um pI de 4,8, por cromatofocalização. A enzima STS apresentou atividade específica de $0,029 \text{ nkatal} \cdot \text{mg}^{-1}$, quando galactinol e rafinose foram utilizadas como substratos. Esta enzima participa tanto da via de síntese de ROs como

de galactosil ciclitóis (Figura 3). A enzima catalisou a reação entre galactinol e D-ononitol formando galactononitol. Ela também transferiu um resíduo galactosil do galactopinitol para rafinose formando estaquiase e D-pinitol (Hoch et al., 1999).

Peterbauer e colaboradores (2002b) encontraram atividade específica de 0,013 nkatal.mg⁻¹ no extrato bruto solúvel das sementes de ervilha para a enzima STS. Os valores de K_m para galactinol e rafinose foram 13,9 mM e 21,1 mM, respectivamente. Esta STS trata-se uma enzima multifuncional transferindo o grupo galactosil inicialmente para a enzima, formando um complexo intermediário enzima-galactose, e posteriormente transferindo a galactose para a rafinose, formando estaquiase (atividade de STS).

Estudos posteriores, realizados por Peterbauer e colaboradores (2003), identificaram um genótipo de ervilha, denominado SD1, apresentando baixo conteúdo de verbascose. O gene *STS* do genótipo SD1 apresentou 99% de identidade com o gene *STS* do cultivar Kelvedon. A enzima STS de ‘SD1’ apresentou atividade específica de estaquiase sintase similar a relatada para a STS do cultivar Kelvedon (Peterbauer et al., 2002b), sendo os valores de K_m determinados para galactinol e rafinose (15,8 mM e 21,0 mM, respectivamente). Entretanto a atividade de verbascose sintase correspondeu a 0,7% da mesma atividade apresentada pela STS de Kelvedon (Peterbauer et al., 2002b). Este fato indica que as mutações na proteína STS de SD1 limitam a capacidade da enzima em usar estaquiase como substrato.

Inan e Keller (2002) descreveram uma enzima que sintetiza estaquiase a partir de duas moléculas de rafinose, denominada galactan:galactosiltransferase (GGT). Esta enzima é capaz de sintetizar estaquiase numa reação independente de galactinol e foi encontrada, até o momento, apenas em vacúolos de folhas em *A. reptans*.

A enzima STS também foi extraída a partir de sementes de soja. A atividade específica da enzima no extrato solúvel das sementes de soja foi 2,15 nkatal.mg⁻¹ e foi detectada apenas na soja madura. Neste estudo a determinação da atividade da enzima STS de soja foi feita apenas na presença dos substratos galactinol e rafinose, impossibilitando a determinação da multifuncionalidade desta enzima (Fialho, 2007).

Uma seqüência nucleotídica que codifica para STS foi isolada pela primeira vez por Peterbauer e colaboradores (1999), a partir de sementes de feijão adzuki. Posteriormente foi isolado um gene que codifica para uma STS multifuncional, que cataliza a síntese de estaquiase e verbascose, em ervilha (Peterbauer et al., 2002b).

As sequências de aminoácidos das STSs de feijão adzuki e de ervilha, codificada pelos genes *VaSTS* e *PsSTS*, respectivamente, apresentam alta similaridade com as sequências codificadas por diferentes genes *RFS* descritos. Tanto a STSs quanto a RFSs reconhecem o galactinol como doador de grupos galactosil e utilizam aceptores que diferem apenas em uma unidade de galactose (Peterbauer et al., 1999; Peterbauer et al., 2002b). Embora STS e RFS apresentem características moleculares e bioquímicas semelhantes, elas correspondem a enzimas distintas em ervilha (Peterbauer et al., 2002b). A STS contém uma inserção central de cerca de 80 aminoácidos, que está ausente em todos os genes *RFS* disponíveis (Peterbauer et al., 1999; Peterbauer et al., 2002b) e é totalmente inativa sobre sacarose (Peterbauer e Richter, 1998; Hoch et al., 1999), enquanto a RFS é inativa sobre a rafinose, como acceptor (Lehle e Tanner, 1973). Além disso, diferentemente da RFS, a STS apresenta ampla especificidade de substratos.

A análise de gel *blot* com o DNA genômico de feijão adzuki, para analisar a família de genes que codificam para estaquiase sintase, sugeriu a presença de dois genes altamente homólogos (Peterbauer et al., 1999). Em *A. meridionalis*, análise de *Southern blot*, sugere que o número de genes desta família é pequeno, sendo que provavelmente não há mais de dois genes *STS* em *A. meridionalis* (Voitsekhovskaja et al., 2009).

2.4. Papel fisiológico dos oligossacarídeos de rafinose

Os ROs representam o segundo grupo de carboidratos solúveis mais abundantes em soja (Minorsky, 2003). Na maioria das espécies vegetais, eles são sintetizados e depositados em órgãos de armazenamento, como sementes e tubérculos, durante o processo de maturação (Dey, 1985; Dey, 1990). Em sementes eles são depositados em todas as partes, incluindo o endosperma, o embrião e a casca (Horbowicz e Obendorf, 1994). Durante os primeiros estágios da germinação, eles são mobilizados pelas α -galactohidrolases (α -D-galactosídeo galactohidrolase, EC 3.2.1.22), servindo como fontes primárias de energia e como substratos para a síntese de outros compostos como os galactosil ciclitolos em diversas espécies de plantas (Dey, 1990; Obendorf et al., 1998; Lahuta et al., 2000; Blöchl et al., 2008). Estudos recentes mostram que, em ervilha e em soja, a inibição da quebra dos ROs, com um inibidor químico de α -galactosidases, durante a embebição e a germinação das sementes, retarda significativamente estes processos (Blöchl et al., 2007; Dierking e Bilyeu, 2009a).

Em plantas, os ROs servem como estoque e fonte de carbono; também podem exercer papéis na tolerância a dessecação durante a maturação das sementes e, em algumas plantas são, juntamente com a sacarose, os principais carboidratos de transporte (Bachmann et al., 1994; Haritatos et al., 1996; Taji et al., 2002; Hannah et al., 2006; Blöchl et al., 2007; Voitsekhovskaja et al., 2009). Acredita-se que uma das funções dos ROs é servir como carboidratos de reserva, transportando energia para os pontos de crescimento dos vegetais (Chatterton et al., 1990).

Em muitas espécies, o acúmulo de ROs coincide com o início da tolerância à dessecação durante os estágios de maturação da semente, por isso estes oligossacarídeos parecem exercer função protetora contra os danos causados durante a dessecação (Horbowicz e Obendorf, 1994). O conteúdo de ROs também foi positivamente correlacionado à longevidade da semente durante o período de armazenamento (Horbowicz e Obendorf, 1994; Lin e Huang, 1994; Haagensohn, 2008).

Os primeiros estágios da dessecação das sementes são marcados pelo acúmulo de sacarose e de ROs, carboidratos que promovem tolerância e previnem possíveis danos causados durante esse processo. A presença de ROs limita a cristalização da sacarose e auxilia na proteção das membranas, proteínas e ácidos nucleicos, durante a desidratação das sementes (Koster e Leopold, 1988; Obendorf, 1997). Os ROs funcionam como osmoprotetores principalmente por serem capazes de manter a integridade das membranas celulares, interagindo com os fosfolípidos de membranas retendo, assim, água durante a dessecação (Bentsink et al., 2000; Zuther et al., 2004).

Os ROs, durante o processo de dessecação, podem formar o estado vítreo, que corresponde a um estado sólido termodinamicamente instável, mas altamente viscoso, que funciona como estabilizante físico e protege contra reações de depreciação intracelulares (Minorsky, 2003). Este estado é considerado como o principal determinante para definição do tempo de armazenagem das sementes (Koster e Leopold, 1988).

A expressão de genes envolvidos na biossíntese de ROs e o acúmulo intracelular destes carboidratos em partes vegetativas das plantas também estão associados com as respostas fisiológicas e bioquímicas a condições de estresses (Taji et al., 2002; Pennycooke et al., 2003; Kaplan et al., 2004; Panikulangara et al., 2004; Peters et al., 2007; Liu et al., 2007; Nishizawa et al., 2008), tal como agentes crioprotetores de órgãos durante a aclimação ao frio (Karner et al., 2004; Zuther et al., 2004). A evidência funcional para o papel dos ROs nos processos de tolerância a estresse

abiótico, em plantas, foi obtido, inicialmente, a partir de plantas transgênicas de *A. thaliana* que acumularam altos níveis de rafinose devido a superexpressão de galactinol sintase (Taji et al., 2002) e através de petúnias transgênicas com atividade de α -galactosidase reduzida (Pennycooke et al., 2003). Estas plantas possuíam elevado conteúdo de rafinose em condições normais de crescimento e mostraram maior tolerância a seca e a exposição a baixas temperaturas, respectivamente, quando comparadas com plantas tipo selvagem.

Embora os ROs exerçam importantes papéis fisiológicos e bioquímicos nos vegetais, em soja, as funções exatas destes oligossacarídeos, durante o desenvolvimento e a germinação das sementes, ainda não foram experimentalmente elucidadas. Estudos recentes demonstraram que não houve diferença significativa entre a germinação de sementes de soja com conteúdo padrão de ROs (selvagem), variedade ‘Williams 82’, e a germinação de sementes de soja com reduzido conteúdo de ROs, quando embebidas/germinadas em água, sugerindo que estes oligossacarídeos não atuam como fonte de energia essencial durante a germinação de semente de soja (Dierking e Bilyeu, 2009a).

2.5. Aspectos nutricionais dos oligossacarídeos de rafinose

Uma das causas que limitam a expansão dos produtos a base de soja no mercado é a presença de fatores antinutricionais nos grãos desta leguminosa. Fatores antinutricionais são moléculas encontradas em uma variedade de alimentos de origem vegetal, que podem provocar resultados adversos ou reduzir a biodisponibilidade de nutrientes quando ingeridos, reduzindo o valor nutritivo destes alimentos e interferindo negativamente no desempenho produtivo e reprodutivo do animal (Santos, 2006). A soja possui diversos fatores antinutricionais, tal como lectinas, inibidores de proteases, ácido fítico, proteínas alergênicas e oligossacarídeos de rafinose (Bajpai, 2005).

Alguns oligossacarídeos, como fructo-oligossacarídeos, podem ser benéficos à saúde, pois ao serem fermentados estimulam o crescimento de algumas bactérias benéficas da microflora intestinal (Kolida et al., 2002). Entretanto, os ROs têm sido descritos como fatores antinutricionais (Kumar et al., 2010). Estes oligossacarídeos podem contribuir para o aumento da viscosidade do bolo alimentar, influenciando a absorção e digestão dos nutrientes uma vez que também interferem com enzimas digestivas no trato intestinal (Smits e Annison, 1996). A presença de concentrações

elevadas destes carboidratos na dieta pode também originar retenção de fluidos, acelerando, conseqüentemente, o fluxo digestivo, afetando diretamente a utilização e absorção dos nutrientes (Wiggins, 1984).

Diversas pesquisas têm estabelecido que os ROs (estaquiose e rafinose) são a principal causa de flatulência associada aos produtos derivados de soja, em humanos e animais (Karlsson et al., 1989; Carlsson et al., 1992; Suarez et al., 1999; Karr-Lilienthal et al., 2005). Os mamíferos são desprovidos de enzimas α -galactosidases, que catalisam a clivagem das reações α -1,6 entre os resíduos galactosil nos ROs. Desta forma, nesses animais, os ROs não são digeridos e nem absorvidos pela mucosa intestinal servindo de substrato para a fermentação realizada pela microflora ativa, principalmente *Clostridium perfringens*, presente no íleo e cólon do intestino grosso (Gitzelmann e Auricchio, 1965; Ruttloff et al., 1967). A fermentação dos ROs resulta na produção de níveis excessivos de gás retal, principalmente CO₂, H₂, CH₄ (Rackis, 1981). A produção de gases pode vir acompanhada de cólicas, diarreia, dispepsia e constipação (Wagner et al., 1976; Suarez et al., 1999; Karr-Lilienthal et al., 2005).

Flatulência é o fator mais importante que determina a freqüência de ingestão de legumes pelas pessoas e pelos animais (De Lumen, 1992). Este fator, induzido por sementes de leguminosas, não compromete apenas os humanos, mas também todos os animais monogástricos. Devido à presença dos ROs, a quantidade de produtos derivados de soja deve ser limitada, por exemplo, em rações para evitar flatulência em cães (*Canis familiaris*) e distúrbios digestivos em filhotes de porcos (*Sus scrofa*) e frangos (*Gallus domesticus*) (Hartwig et al., 1997).

Algumas pesquisas têm demonstrado que a ingestão de produtos derivados de soja sem ou com baixos teores destes oligossacarídeos solúveis promove resultados nutricionais benéficos, incluindo redução da viscosidade do bolo alimentar e aumento da absorção de nutrientes (Coon et al., 1990; Leske et al., 1991; Parsons et al., 2000; Smiricky et al., 2002). Segundo Suarez e colaboradores (1999) humanos que consumiram farinha de soja derivada de sementes com baixos teores de ROs, apresentaram redução nos níveis de flatulência. Brasil (2007) demonstrou que houve melhora significativa da digestibilidade em ratos tratados com dieta contendo farinha de soja sem ROs quando comparados a ratos submetidos a dietas de farinha com ROs.

Com o objetivo de atenuar os efeitos provocados por estes fatores antinutricionais, vários métodos, que visam à redução ou a eliminação dos ROs, têm sido descritos na literatura: hidrólise enzimática (Guimarães et al., 2001), hidratação

dos grãos (Sanni et al., 1997), cozimento (Pugalenthi et al., 2006), fermentação (Rehms e Braz, 1995; Donkor et al., 2007) e extração dos ROs com água (Ku et al., 1976) e com etanol (Martínez-Villaluenga et al., 2006).

Embora alguns destes métodos tenham proporcionado resultados bastante satisfatórios em reduzir a ação dos ROs como fatores antinutricionais, uma estratégia mais direta seria a redução dos ROs nas sementes de soja, através da manipulação genética e bloqueio da atividade das enzimas da via de síntese destes oligossacarídeos. Desta forma, a obtenção de variedades de soja com reduzido conteúdo de ROs seria possível.

Segundo Jones e colaboradores (1999), uma estratégia adequada para a obtenção de genótipos de soja com baixo teor de ROs seria a manipulação genética de cada etapa individual da via de biossíntese. A RFS é a enzima que conduz sacarose para a via dos oligossacarídeos de rafinose (Peterbauer et al., 2002a), e, portanto, representa um ponto de controle a partir do qual a manipulação poderia evitar a síntese dos ROs. Porém, os oligossacarídeos têm grande importância fisiológica, eliminá-los completamente poderia resultar em graves consequências para as plantas. Alternativamente, a manipulação genética da próxima enzima envolvida na via biossintética, a STS, pode proporcionar a obtenção de variedade de soja, não totalmente isenta dos ROs, mas com baixo teor, o que reduziria os problemas de flatulência, e ampliaria a aceitação e o consumo de tais produtos.

2.6. Variedades de soja com reduzido conteúdo de ácido fítico e oligossacarídeos de rafinose

Em sementes de soja, a maioria do fósforo inorgânico está na forma de ácido fítico ou fitato (mio-inositol 1, 2, 3, 4, 5, 6-hexafosfato). O fósforo presente no ácido fítico não está nutricionalmente disponível e não é absorvido por animais não-ruminantes uma vez que estes não possuem a enzima fitase, que age sobre o fitato hidrolisando-o (Erdman, 1979; Raboy, 1997). O ácido fítico também é um forte agente quelante que se complexa com íons metálicos, reduzindo a quantidade de íons ferro, zinco, cálcio e magnésio em produtos derivados de soja (Brown e Solomons, 1991; Mendoza et al., 1998). Assim, visando aumentar disponibilidade de fósforo e a qualidade nutricional destes produtos comerciais, pesquisas têm sido realizadas para proporcionar a obtenção de variedades de soja com baixo conteúdo de fitato.

O ácido fítico é sintetizado a partir da fosforilação do mio-inositol, o qual também funciona como substrato para biossíntese de ROs (Figura 4). Assim, a busca por variedades de soja com reduzido conteúdo de fitato resultou também na obtenção de sementes e soja com baixo conteúdo de ROs. Sebastian e colaboradores (2000), através de mutagênese química e melhoramento genético, identificaram uma planta de soja, PI 200508, com reduzido conteúdo de fitato e de ROs, mas elevada quantidade de sacarose. A diminuição do conteúdo de fitato nesta planta foi ocasionada por um alelo recessivo mutante do gene *MIPS*, que codifica para mio-inositol-1-fosfato sintase (MIPS), denominado *mips*. As linhagens de soja com genótipo *mips mips* apresentaram um conteúdo de rafinose e estaquiose de 10 e 5 $\mu\text{mol.g}^{-1}$, respectivamente, enquanto as linhagens de soja convencionais, contendo o genótipo *Mips Mips*, apresentaram um conteúdo de rafinose e estaquiose de 20 e 75 $\mu\text{mol.g}^{-1}$, respectivamente.

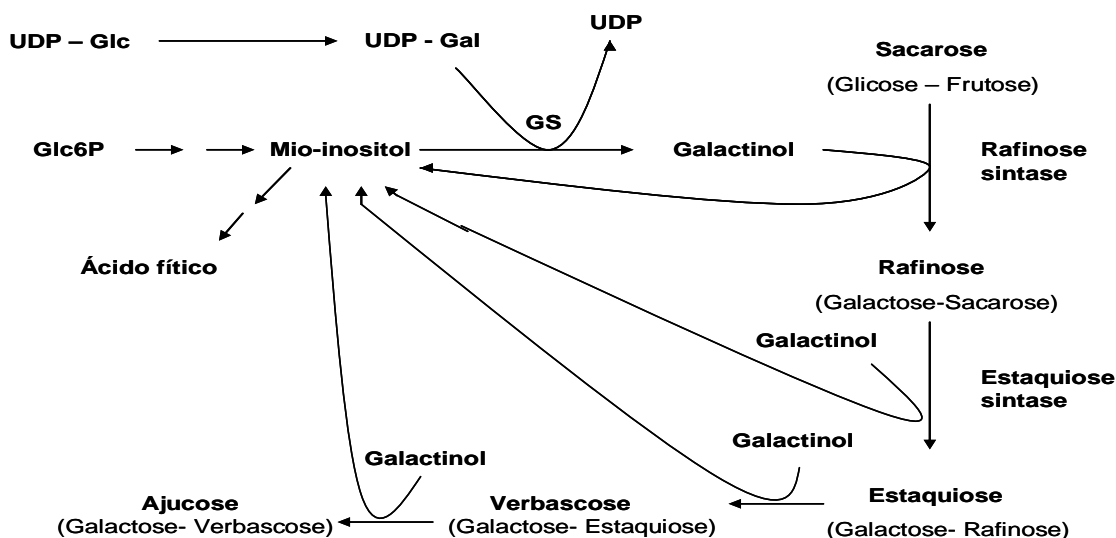


Figura 4 Representação esquemática das vias de síntese de oligossacarídeos de rafinose e ácido fítico em soja. Mio-inositol funciona como substrato comum para as duas vias metabólicas. UDP-Glc, uridina difosfato glicose; UDP-Gal, uridina difosfato galactose; Glc6P, glicose-6-fosfato. (Adaptado de Suarez et al., 1999).

Segundo Hitz e colaboradores (2002), sementes de soja submetidas à mutação induzida no gene *MIPS*, apresentaram fenótipo com elevado nível de fosfato inorgânico, e, reduzido conteúdo de fitato e ROs. A mutação induzida correspondeu a uma substituição de um resíduo do aminoácido lisina 396 para um resíduo de asparagina. Essa troca resultou no decréscimo da atividade específica da enzima mio-inositol-1-fosfato sintase de aproximadamente 90% e, conseqüentemente a redução de mio-inositol-1-fosfato. O novo genótipo, com reduzido conteúdo de ROs, foi posteriormente

analisado quanto ao vigor da muda e viabilidade da semente (Hitz et al., 2004; Neus et al., 2005).

Nunes e colaboradores (2006), através do silenciamento do gene *GmMIPSI* (que codifica para mio-inositol-1-fosfato sintase), obtiveram linhagens de soja transgênicas nas quais os níveis de fitato e o desenvolvimento das sementes foi reduzido, sugerindo uma importante correlação entre a expressão do gene *GmMIPSI* e o desenvolvimento de sementes. Nestas linhagens, produzidas pelo uso da técnica RNA de interferência (RNAi), o gene *GmMIPSI* demonstrou estar silenciado pois não se detectou a presença dos transcritos *GmMIPSI*. Uma drástica redução do conteúdo de fitato foi alcançada nestas linhagens transgênicas (94,5%), entretanto não foram realizadas análises do conteúdo de ROs.

O efeito da mutação no gene *MIPSI*, em um novo mutante de soja, Gm-Ipa-TW-1, com reduzido conteúdo de fitato, foi avaliado quanto à qualidade da semente e suas características nutricionais. Análises demonstram que a mutação não afetou a concentração protéica, o conteúdo de aminoácidos e de ácidos graxos individuais. Entretanto, este mutante apresentou variação no nível de carboidratos solúveis: houve aumento no teor de sacarose (47,4-86,1%) e diminuição significativa no conteúdo de rafinose (74,2-84,3%) (Yuan et al., 2009), quando comparadas às plantas controles. Este estudo indicou que mutações nos genes envolvidos na biossíntese de carboidratos não afetam negativamente a qualidade nutricional da semente de soja. Ao invés disto, tem o potencial de melhorar algumas de suas propriedades.

Segundo Sebatian e colaboradores (2000), a redução do teor de fitato na linhagem PI 200508 foi devido ao alelo *mips*; entretanto a redução de rafinose e estaquiose foi determinada por outro alelo recessivo mutante, denominado *stc1*. O termo ‘locus *Stc1*’ se refere a um locus que afeta o conteúdo de ROs em sementes de soja. Sementes de soja homozigotas para o alelo mutante *stc1a* apresentam baixo conteúdo de rafinose e estaquiose, mas alto teor de galactinol e sacarose (Obendorf et al., 2008).

Segundo Dierking e Bilyeu (2008), o genoma da soja PI 200508 contém pelo menos dois genes *GmRFS* expressos que são similares a outros genes *RFS* caracterizados, sendo que mutações em um deles, no *GmRFS2*, resultaram no alelo *stc1*. Nesta linhagem mutante, o baixo teor de rafinose é causado pela baixa atividade da enzima GmRFS2. Mutações no gene *GmRFS2* resultam na alteração de dois aminoácidos de interesse. A primeira é a substituição de um resíduo de aminoácido cisteína 171 para um resíduo de serina. A segunda corresponde a uma deleção de um

resíduo e triptofano 331 em uma região altamente conservada. Embora uma destas alterações resulte no fenótipo de reduzido conteúdo de ROs, não se pode determinar qual delas é efetivamente responsável, pois o domínio catalítico desta enzima não foi ainda determinado.

O baixo conteúdo de rafinose, nesta planta, resulta no reduzido acúmulo de estaquiose, mesmo na presença de STS e GolS ativas (Hitz et al., 2002), e não altera o acúmulo de galactosil ciclitolis, os quais têm demonstrados funções similares as dos ROs durante a maturação e na tolerância à dessecação em sementes de soja (Obendorf et al., 1997).

O alelo recessivo *stc1* foi introduzido em variedades de soja comerciais utilizadas como fonte protéica. E embora os cultivares com o alelo *stc1* apresentem uma quantidade significativamente reduzida de rafinose e estaquiose, e, elevado nível de sacarose e galactinol, eles demonstram um rendimento menor que os cultivares convencionais contendo o alelo *Stc1*. Entretanto, a redução destes açúcares não demonstrou afetar a morfologia das folhas, o tamanho da planta e as características das sementes (Neus et al., 2005).

Schillinge e colaboradores (2011) identificaram um novo mutante de soja, originado do mutante PI200508, denominado SG-ULRFO. Este mutante apresenta um conteúdo de rafinose e estaquiose muito baixo (*ultra-low*). A análise das sequências dos genes *GmRFS* presentes neste mutante, revelaram que alterações nos genes *GmRFS3* e *GmRFS4* resultam em grande redução do conteúdo de rafinose e estaquiose no mesmo. O conteúdo de rafinose foi inferior a 0,13% da massa seca total, enquanto o teor de estaquiose foi menor que 1,6% da massa seca total da semente. O mutante SG-ULRFO também demonstrou aumento na taxa de germinação das sementes (83%) quando comparadas com sementes de soja que não tem um teor de rafinose e estaquiose muito baixos.

2.7. Silenciamento gênico pós-transcricional por RNA

O silenciamento gênico representa uma via de controle da expressão gênica que consiste na repressão da expressão de um gene alvo pela destruição do transcrito ou pela inibição da transcrição, baseando-se na interação entre sequências de ácido ribonucléico (RNA) (Tang e Galili, 2004).

O silenciamento gênico por meio de RNA é um mecanismo celular natural e parece desempenhar, primordialmente, um papel na eliminação de RNAs mensageiros anômalos, no remodelamento da cromatina e na defesa do organismo contra transposons e vírus (Tabara et al., 1999; Ketting et al., 1999; Baulcombe, 2004; Almeida e Allshire, 2005). Os procariotos não apresentam esse mecanismo, considerado, portanto, uma aquisição ou inovação eucariótica (Zamore, 2002).

O silenciamento por RNA não afeta a transcrição de um locus gênico, mas somente causa degradação de sequências específicas de mRNAs alvos. Neste sentido, um RNA dupla fita (dsRNA – *double stranded RNA*) é clivado em pequenos RNAs, siRNA (*small interfering*), para se tornar funcional em um número de processos de silenciamento gênico (Yu e Kumar, 2003). Estes pequenos RNAs são reguladores ativos da expressão gênica, atuando de forma complexa e coordenada (Silva et al., 2004).

Das metodologias de silenciamento gênico induzido por dsRNA, a interferência por RNA (*RNA interference*) é a mais eficiente (Smith, 2000). Essa metodologia é mais eficiente em termos de extensão do gene silenciado e o silenciamento resultante é quase tão completo quanto o proporcionado pelo método *knockout* (Tang et al., 2007).

O processo de RNAi foi descoberto como um fenômeno induzido em nematóides (Hannon, 2002). A evidência inicial do silenciamento gênico por RNAi foi relatada em *Caenorhabditis elegans* através do uso de RNA *antisense* (Guo e Kemphues, 1995). Neste estudo, Guo e Kemphues demonstraram que a introdução da fita de RNA *antisense* foi capaz de bloquear a expressão do gene *par-1*, responsável por estabelecer a polaridade e a divisão assimétrica de embriões de *C. elegans* e suprimir a função do gene de forma sequência-específica, garantindo a diversidade das células durante o desenvolvimento. Em estudo posterior, Fire e colaboradores (1998) introduziram dsRNA (fita *sense e antisense*) em *C. elegans* e verificaram silenciamento muito mais eficiente do que quando foi utilizado cada fita isoladamente e que poucas moléculas eram necessárias para silenciar completamente a expressão do gene homólogo (Mahmood-ur-Rahaman et al., 2008).

Durante a última década, pesquisas demonstraram o efeito do RNAi em mamíferos (Hannon, 2002; Silva et al., 2004; Angaji et al., 2010), em insetos (Kennerdell e Carthew, 2000; Kavi et al., 2008; Mueller et al., 2010), em anfíbios (Dirks et al., 2003; Li e Rohrer, 2006), e em plantas (Nunes et al., 2006; Hirai e Kodam, 2008; Flores et al., 2008; Watanabe, et al. 2010; Runo et al., 2010). Em plantas, o silenciamento gênico via RNAi desempenha importante papel modulando a maioria dos

processos celulares importantes, tais como crescimento, desenvolvimento, além de estar associado a respostas bioquímicas à estresses (López-Gomollón e Dalmay, 2011), inclusive a estresses bióticos (Senthil-Kumar e Mysore, 2011).

O processo base envolve um dsRNA, com seqüência similar ou perfeitamente igual a do gene a ser silenciado, que é processado em pequenas unidades que guiam o reconhecimento e a clivagem e/ou a inibição da tradução de mRNA alvos homólogos. O dsRNA, que ativa a via RNAi, pode ser originado no núcleo ou no citoplasma de vários modos, incluindo transcrição por seqüências repetidas e invertidas de DNA, simultânea síntese de RNA *sense* e *antisense*, replicação viral, e atividade da RNA polimerase dependente de RNA (RdRP) sob molde de RNA fita simples (Matzke et al., 2001).

O mecanismo do silenciamento por RNA induzido por dsRNA pode ser resumido de forma simples como possuindo dois maiores passos: o passo de iniciação e o passo efetor (Cerutti, 2003). O passo de iniciação envolve a enzima Dicer, DCR ou DCL (*Dicer-like*) (Jones-Rhoades et al., 2006; Dunoyer et al., 2010) que reconhece e cliva um dsRNA alvo em siRNAs de aproximadamente 21 – 26 nt com uma projeção de 2 nt na extremidade 3', que corresponde à ambas as fitas *sense* e *antisense* (Perrimon et al., 2010). A Dicer é uma proteína altamente conservada, que apresenta dois domínios modulares de RNase III e é específica para dsRNA (Zamore et al., 2000; Runo et al., 2010).

O passo efetor ocorre no citossol e envolve o complexo ribonucleoprotéico conhecido com complexo de silenciamento induzido por RNA (RISC - *RNA-induced silencing complex*). Este complexo efetor possui siRNA, mRNA complementares, proteínas argonautas (AGO), as quais contêm dois domínios - MID/Piwi (o motivo MID se liga a extremidade 5' do siRNA guia e o motivo Piwi tem atividade de ribonuclease) e PAZ (este domínio se liga ao terminal 3' do siRNA guia) (Baumberger e Baulcombe, 2005; Qi et al., 2005; Guildiyal e Zamore, 2009; Riedmann e Schwentner, 2010), além de diversos outros fatores que contribuem para sua funcionalidade (Hannon, 2002). Neste passo os siRNAs fita simples são incorporados ao complexo RISC no qual a degradação ou a inibição da tradução do mRNA alvo ocorre com os pequenos RNAs agindo como guia. Cada RISC parece ter um único siRNA, uma RNase e um domínio de reconhecimento e ligação do mRNA homólogo. Cada clivagem de mRNA alvo ocorre freqüentemente pelo centro do siRNA dentro do complexo RISC, resultando em mais de 90% de inibição da expressão gênica (Cerutti, 2003; Yu e Kumar, 2003).

2.7.1. Técnicas para indução do silenciamento gênico via RNAi

Embora o silenciamento gênico induzido por dsRNA, em plantas, tenha sido descoberto por Waterhouse e colaboradores (1998), a primeira aplicação desta tecnologia foi realizada por Chuang e Meyerowitz (2000). Neste caso, foram construídos vetores de transformação que produziam RNA capazes de formar duplex (dsRNAs) com sequências iguais a dos genes alvos, ligadas na orientação *sense* e *antisense*, sob controle de um promotor viral forte. Estes dsRNAs engatilharam o silenciamento eficiente de quatro genes em *A. thaliana*.

Os vetores RNAi foram projetados especificamente para gerar espécies de dsRNAs longas com a mesma seqüência do gene alvo. Esta tecnologia de RNAi, baseada em vetores, foi simultaneamente melhorada pela utilização de um íntron, como ligante, para formação de um RNA *hairpin* (hpRNA) (Smith, 2000). Os vetores fabricados para expressar hpRNA foram também empregados com sucesso para promover o silenciamento de genes alvos correspondentes (Hirai e Kodama et al., 2008).

Entretanto, em muitos casos, a expressão constitutiva de dsRNA conduz a efeitos adversos inesperados para o crescimento e desenvolvimento da planta. Para solucionar esta questão, vetores de RNAi induzidos quimicamente tem sido desenvolvidos, sendo capazes de controlar temporal e espacialmente o silenciamento da expressão gênica. Um sistema RNAi induzível ideal deve ser rapidamente induzido em toda planta ou em um tecido específico, além de poder ser inativado em um estado não induzível para impedir o silenciamento gênico indesejável (Wielopolska et al., 2005).

O controle tecido-específico ou órgão-específico tem demonstrado uma alternativa eficiente para promover o silenciamento via RNAi no melhoramento vegetal (Tang et al., 2007). A combinação da expressão de genes alvos e promotores órgão/tecido-específicos regulados por RNAi possibilita destruição espacial limitada da expressão do mRNA alvo e é útil para determinar ou confirmar a função de um gene, além de possibilitar o melhoramento dos produtos derivados de plantas (Islam et al., 2005)

Atualmente existem três métodos para introduzir DNA em células vegetais objetivando silenciamento gênico via RNA: bombardeamento de partícula, infecção por *Agrobacterium* e infecção por vírus (Watson et al., 2005). O bombardeamento de células vegetais com dsRNA, siRNA ou construções de DNA que codificam hpRNA

podem produzir silenciamento gênico estável ou transiente nas células transformadas. De forma similar, as construções de DNA que codificam esses tipos de RNA, quando colocadas em um T-DNA de *Agrobacterium* e usadas para transformar células vegetais, induzem o silenciamento gênico de genes alvos. Uma alternativa para introdução de dsRNA em plantas é o uso de vírus (*virus induced gene silencing* – VIGS) (Silva et al., 2004). A infecção de uma planta com um vírus recombinante, que contém uma seqüência de um gene endógeno nuclear para a planta, pode resultar no silenciamento daquele gene. O sistema não conta com a transformação da planta, apesar de exigir a clonagem de seqüências do gene alvo dentro do genoma viral recombinante (Wang e Waterhouse, 2001).

2.7.2. Principais avanços do uso do RNAi no melhoramento de plantas

Os primeiros avanços do silenciamento via RNAi foram conseguidos no desenvolvimento de plantas resistentes a doenças, particularmente aquelas causadas por vírus, considerando que este tipo de silenciamento é um mecanismo natural de resistência aos vírus em planta (Soosaar et al., 2005). Segundo Parrott e colaboradores (2010), a geração atual de plantas resistentes a vírus tem sido gerada através do uso de construções que resultam em hpRNAs.

A efetiva utilização da tecnologia de RNAi para gerar plantas resistentes a vírus foi primeiramente obtida por Waterhouse e colaboradores (1998), que desenvolveram imunidade em plantas de batata ao *Potato vírus Y* (vírus de RNA, família *Potyviridae*), através da expressão *sense* e *antisense* do transcrito do gene *HC-Pro*. Em 2003, foram alcançadas plantas resistentes a outras viroses via RNAi: plantas de *V. mungo*, que se recuperavam de infecção por *Mungbean yellow mosaic indiovirus* (vírus de DNA, família *Geminiviridae*) (Pooggin et al., 2003); redução do acúmulo do *African cassava mosaic vírus* (vírus de DNA, família *Geminiviridae*) em protoplastos de tabaco (Vanitharani et al., 2003); bloqueio da infectividade do *Pepper mild mottle vírus* (vírus de RNA, família *Tobamoviridae*) em plantas de tabaco (Tenllado et al., 2003); o não aparecimento de sintomas característicos do *Tobacco etch vírus* (vírus de RNA, família *Potyviridae*) em plantas de tabaco (Tenllado et al., 2003); recuperação da infecção ao *Alfafa mosaic vírus* (vírus de RNA, família *Bromoviridae*) em plantas de tabaco (Tenllado et al., 2003). No trabalho de Andika e colaboradores (2005) foram obtidas plantas de tabaco tolerantes ao *Beet necrotic yellow vein vírus* (vírus de RNA, família

Benyviridae). Em estudos posteriores, Zhao e colaboradores (2006), através da expressão transiente de um pequeno hpRNA da proteína de 126 kDa do *Tobacco mosaic virus* -TMV (vírus de RNA, família *Tobamoviridae*), inibiu a replicação deste vírus em plantas de tabaco.

A utilização da tecnologia de RNAi possibilitou que várias regiões do genoma viral fossem alvos do processo de silenciamento gênico em plantas que expressam sequências virais como dsRNA ou hpRNA. Segundo Runo e colaboradores (2010) diversos segmentos do genoma viral podem fornecer resistência estável e duradoura, embora o genoma viral seja altamente recombinante.

Outro interessante avanço na utilização do RNAi, é a produção de culturas resistentes a insetos e nematóides. Plantas transgênicas de milho expressando dsRNA, que tinham como alvo a enzima ATPase vacuolar (*V-type ATPase A*) de células do intestino de crisomelídeos (coleópteros que atacam o milho), demonstraram uma redução dos danos causados por estes insetos, como consequência da diminuição da taxa de crescimento e aumento da mortalidade larval dos mesmos (Baum et al., 2007).

Para resistência a nematóides, o uso desta tecnologia tem sido descrita em plantas de tabaco, transformadas via *Agrobacterium*, expressando dsRNA contra dois genes, um codifica para um fator envolvido em *splicing* e outro que codifica para uma proteína integrase, do nematóide *Meloidogyne incognita*. Estas plantas apresentaram resistência ao nematóide, dando suporte a teoria de que a produção de dsRNA pelo hospedeiro pode ser uma estratégia eficiente para controle de doenças parasitárias em plantas (Yadav et al., 2006).

Recentemente, raízes de soja foram transformadas, via *Agrobacterium rhizogenes*, com construções contendo dois genes de *M. incognita*, um que codifica para tirosina fosfatase e outro que codifica para um precursor de uma proteína mitocondrial, visando o silenciamento dos mesmos, via RNAi. Estes dois genes apresentam alta similaridade com os referidos genes do nematóide do cisto da soja, *Heterodera glycines*. Como resultado, observou-se forte redução (92% e 94,7%, respectivamente) da formação de galhas durante a infecção por *M. incognita*, demonstrando que esta metodologia pode ser utilizada para o desenvolvimento de variedades de soja com resistência a nematóides (Ibrahim et al., 2011).

Além disso, existem circunstâncias onde há potencial utilização da técnica de RNAi, como: redução dos níveis de expressão de uma neurotoxina denominada β -oxalilaminoalanina-L-alanina (BOAA), encontrada na leguminosa *Lathyrus sativus*, a

qual está localizada em países como Etiópia, Bangladesh e Índia. Nestes países, em situações de inanição, estas plantas são ingeridas provocando paralisia temporária ou permanente dos membros (Willians et al., 2004).

A utilização da RNAi para elevar especificamente o valor nutricional dos órgãos vegetais que são consumidos como alimentos está aumentando progressivamente, com vários esforços bem sucedidos sendo publicados, sugerindo que esta metodologia tem potencial enorme para o melhoramento vegetal (Tang et al., 2007). Neste sentido, o silenciamento da expressão de genes específicos resulta em plantas com características desejáveis (Vaucheret et al., 2001; Tang e Galili, 2004; Mansoor et al., 2006).

Contribuições foram realizadas via RNAi para o melhoramento das plantas de arroz (Kusaba et al., 2003). Foi possível diminuir o nível, via hpRNA, de gluteína (principal proteína de armazenamento na semente), uma vez que pacientes com problemas de rins não podiam digerir tal proteína. Foi produzida uma variedade denominada LGC-1 (*low glutenin content 1*) e este caráter foi estável por 20 gerações, demonstrando a estabilidade do dsRNA em plantas transgênicas.

A tecnologia de RNAi também tem sido usada para promover o melhoramento de plantas de café. Neste caso, o conteúdo de cafeína foi reduzido significativamente em plantas transgênicas de café por silenciamento do gene que codifica para a enzima cafeína sintase (Ogita et al., 2004).

O uso de hpRNA para silenciar o gene que codifica a enzima 1-aminociclo propano-1-carboxilato oxidase (ACC oxidase), também resultou em plantas transgênicas de tomate com sensibilidade a etileno reduzida, aumentando assim o tempo de vida de prateleira de seus frutos (Xiong et al., 2005). Além disso, em tomate, uma construção hpRNA foi empregada para suprimir um gene regulatório da fotomorfogênese endógena, *DET1*, usando um promotor fruto específico. Plantas transgênicas demonstraram um aumento nas quantidades de carotenóides e flavonóides, comparado com plantas não transformadas, enquanto todos os outros parâmetros de qualidade do fruto não foram alterados (Davuluri et al., 2005).

A produção e a qualidade das sementes de um mutante de milho rico em lisina foi melhorada via RNAi. Neste mutante, com alto conteúdo de resíduos de lisina, genes que codificam para a proteína de reserva 22 kDa, que é abundante em milho, mas com baixo conteúdo de lisina, foram silenciados via RNAi, usando um promotor semente específico. A regulação negativa dos genes resultou em sementes de milho normais e de boa qualidade, com altos teores deste aminoácido (Segal et al., 2003).

Sunilkumar e colaboradores (2006), objetivando reutilizar os resíduos de sementes de algodão após extração da fibra, promoveram a redução da concentração de gossipol nas sementes por silenciamento via RNAi. A semente de algodão contém gossipol, um terpenóide tóxico, que impede a utilização como fonte de proteínas e calorias. Para eliminá-lo, plantas de algodão foram transformadas com a construção para silenciamento do gene que codifica σ -cadieno sintase, fusionado a um promotor semente específico. Estas plantas tiveram o teor de gossipol em suas sementes reduzidos, enquanto o conteúdo deste terpenóide nos demais tecidos foi comparável aos das plantas controle.

Outra grande aplicação da técnica RNAi tem sido a modificação genética do teor de ácidos graxos em óleos vegetais. Por exemplo, o silenciamento de duas enzimas chave na via de biossíntese de ácidos graxos, *ghSAD-1* e *ghFAD2-1*, modificou a composição destes ácidos em algodão. A subregulação, mediada por RNAi, de *ghSAD-1* elevou o conteúdo de ácido esteárico em sementes de algodão (44% comparado com o nível normal de 2%), e o silenciamento de *ghFAD2-1* aumentou o conteúdo de ácido oléico (77% comparado com o nível normal de 15%) (Liu et al., 2002).

Para o melhoramento da qualidade nutricional da soja foi conduzida uma transformação via *Agrobacterium* para promover o silenciamento do gene *GmFAD2-1* (gene da oleoil dessaturase) via RNAi. Uma construção contendo uma cópia invertida do fragmento do gene *GmFAD2-1*, espaçada por um íntron, sob controle do promotor da lectina (semente específico) foi usada e foram obtidas linhagens com alto teor de ácido oléico: 71,5-81,9 % nas linhagens transformadas comparado a 18,1% na semente de soja parental (Wang e Xu, 2008).

Flores e colaboradores (2008) também demonstraram o silenciamento gênico mediado por hpRNA do gene do ácido graxo Omega-3-dessaturase (*FAD3*) em soja. A enzima FAD3 é responsável pela síntese de ácidos α -linolênico na via de ácidos graxos poliinsaturados. Os autores demonstraram que siRNAs, gerados a partir de um transgene contendo repetições invertidas de um domínio conservado entre os membros desta família gênica, são efetivos em silenciar o gene *FAD3* em plantas de soja engenheiradas, resultando em um nível muito baixo de ácidos graxos α -linolênicos.

Este trabalho também mostrou que tal silenciamento é estavelmente herdado nas linhagens de soja engenheiradas. Estes resultados sugerem um papel e um grande potencial para hpRNA em processos de silenciamento gênico de famílias multigênicas presentes em genomas complexos, tal como a soja.

3. REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida R, Allshire RC (2005) RNA silencing and genome regulation. *Trends in Cell Biology* 15: 251-258.
- Andika IB, Kondo H, Tamada, T (2005) Evidence that RNA silencing-mediated resistance to beet necrotic yellow vein virus is less effective in roots than in leaves. *Molecular Plant-Microbe Interactions* 18: 194–204.
- Angaji SA, Hedayati SS, Poor RH, Madani S, Poor SS, PANAHI, S (2010) Application of RNA interference in treating human diseases. *Journal of Genetics* 89: 527-537.
- Bachmann M, Keller F (1995) Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. Inter- and intracellular compartmentation. *Plant Physiology* 109: 991–998.
- Bachmann M, Matile P, Keller F (1994) Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. (cold acclimation, translocation, and sink to source transition: discovery of chain elongation enzyme). *Plant Physiology* 105: 1335–1345.
- Bajpai S, Sharma A, Gupta MN (2005) Removal and recovery of antinutritional factors from soybean flour. *Food Chemistry* 89: 497-501.
- Baulcombe D (2004) RNA silencing in plants. *Nature* 431: 356-363.
- Baum JA, Bogaert T, Clinton W, Heck GR, Feldmann P, Ilagan O, Johnson S, Plaetinck G, Munyikwa T, Pleau M, Vaughn T, Roberts J (2007) Control of coleopteran insect pests through RNA interference. *Nature Biotechnology* 25: 1322-1326.
- Baumberger N, Baulcombe DC (2005) *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences of the United States of America* 102: 11928–11933.

- Beebe DU, Turgeon R (1992) Localization of galactinol, raffinose, and stachyose synthesis in *Cucurbita pepo* leaves. *Planta* 188: 354-361.
- Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier K, Groot SPC, Koornneef M (2000) Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiology* 124:1595–1604.
- Blöchl A, Peterbauer T, Hofmann J, Richter A (2008) Enzymatic breakdown of raffinose oligosaccharides in pea seeds. *Planta* 228: 99–110.
- Blöchl A, Peterbauer T, Richter A (2007) Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. *Journal of Plant Physiology* 164: 1093-1096.
- Brasil APR (2007) Avaliação bioquímica e nutricional de farinha de soja processada enzimaticamente para remoção dos oligossacarídeos de rafinose. Universidade Federal de Viçosa, Viçosa-MG, p.90. (Dissertação M.S.).
- Brown, KH, Solomons NW (1991) Nutritional problems of developing countries. *Infectious Disease Clinics of North America* 5: 297–317.
- Carlsson NG, Karlsson H, Sandberg ASJ (1992) Determination of oligosaccharides in foods, diets, and intestinal contents by high-temperature gas chromatography and gas chromatography/mass spectrometry. *Journal of Agricultural and Food Chemistry* 40: 2404–2412.
- Castillo EM, De Lumen BO, Reyes PS, De HZ (1990) Raffinose synthase and galactinol synthase in developing seeds and leaves of legumes. *Journal of Agricultural and Food Chemistry* 38: 351–355.
- Cerutti H (2003) RNA interference: traveling in the cell and gaining functions? *Trends in Genetics Review* 19: 39-45.
- Chatterton NJ, Harrison PA, Thornley WR, Draper EA (1990) Oligosaccharides in foliage of *Agropyron*, *Bromus*, *Dactylis*, *Festuca*, *Lolium* and *Phleum*. *New Phytologist* 114: 167-171.
- Chuang CF, Meyerowitz EM (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the USA* 97: 4985–4990.
- Companhia Nacional de Abastecimento (CONAB) (2011a) Acompanhamento da safra brasileira – Grãos – Primeiro levantamento safras 2010/2011. Disponível em: http://www.conab.gov.br/OlalaCMS/uploads/arquivos/11_01_06_08_41_56_boletim_graos_4o_lev_safra_2010_2011..pdf. Acesso em 12/02/2011.

- Companhia Nacional de Abastecimento (CONAB) (2011b) Balança do Agronegócio exportações brasileiras. Disponível em: http://www.conab.gov.br/OlalaCMS/uploads/arquivos/11_01_12_16_41_30_0205_balanca__exportacao..pdf. Acesso em 12/02/2011.
- Coon CN, Leske KL, Akavanichan O, Chen TK (1990) Effect of oligosaccharides-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. *Poultry Science* 69: 787-793.
- Davuluri GR, Van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J, Bramley PM, Pennings HM, Bowler C (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nature Biotechnology* 23: 890–895.
- De Lumen BO (1992) Molecular strategies to improve protein quality and reduces flatulence in legumes: A review. *Food Structure* 11: 33-46.
- Deleprane JB, Pacheco J, Boavenura GT (2009) Evaluation of protein quality from genetically modified and organic soybean in two consecutive generations of wistar rats. *Brazilian Archives of Biology and Technology* 52: 841-847.
- Dey PM (1985) D-galactoside containing oligosaccharides. In: *Biochemistry of Storage Carbohydrates in Green Plants*, p. 53-129.
- Dey PM (1990) In: *Methods in plant biochemistry*. London: Academic Press, p. 189-218.
- Dierking EC, Bilyeu KD (2008) Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 2: 135-145.
- Dierking EC, Bilyeu KD (2009a) Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *Journal of Plant Physiology* 166: 1329-1335.
- Dierking EC, Bilyeu KD (2009b) New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biology* 9:89.
- Dirks RP, Bouw GB, Huizen RR, Jansen EJ, Martens JM (2003) Functional genomics in *Xenopus laevis*: Towards transgene-driven RNA interference and cell-specific transgene expression. *Current Genomics* 4: 699–711.
- Downie B, Gurusinghe S, Dahal P, Thacker RR, Snyder JC, Nonogaki H, Yim K, Fukunaga K, Alvarado V, Bradford KJ (2003) Expression of a galactinol synthase gene in tomato seeds is up-regulated before maturation desiccation and again after

- imbibition whenever radicle protrusion is prevented. *Plant Physiology* 131: 1347–1359.
- Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, Alioua A, Himer C, Voinnet O. (2010) An endogenous, systemic RNAi pathway in plants. *EMBO Journal* 29: 1699-1712.
- Egounlety M, Aworh OC (2003) Effect of soaking, dehulling, cooking and fermentation with *Rhizopus oligosporus* on the oligosaccharides, trypsin inhibitor, phytic acid and tannins of soybean (*Glycine max* Merr.), cowpea (*Vigna unguiculata* L. Walp) and groundbean (*Macrotyloma geocarpa* Harms). *Journal of Food Engineering* 56: 249-254.
- Erdman JW (1979) Oilseed phytates: nutritional implications. *Journal of the American Oil Chemists' Society* 56: 736–741.
- Esteves EA, Monteiro JBR (2001) Efeitos benéficos das isoflavonas de soja em doenças crônicas. *Brazilian Journal of Nutrition* 14: 43-52.
- Fialho L da S (2007) Purificação e caracterização de uma α -galactosidase em sementes de *Tachigali multijuga* e clonagem parcial do gene da estaquiose sintase de soja. Universidade Federal de Viçosa – Viçosa, MG, p. 140. (Tese D.S.).
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806-11.
- Flora LL, Madore MA (1993) Stachyose and mannitol transport in olive (*Olea europaea* L.). *Planta* 189: 484-490.
- Flores T, Karpova O, Su X, Zeng P, Bilyeu K, Sleper DA, Nguyen HT, Zhang ZJ (2008) Silencing of GmFAD3 gene by siRNA leads to low α -linolenic acids (18:3) of fad3-mutant phenotype in soybean [*Glycine max* (Merr.)]. *Transgenic Research* 17: 839-850.
- Food and Agricultural Organization of the United Nations (FAO) (2010) FAO statistical tables, agriculture [Online]. Disponível em: <http://faostat.fao.org/site/339/default.aspx>. Acesso em: 12/02/2011.
- Frank T, Nörenberg S, Engel KH (2009) Metabolite profiling of two novel low phytic acid (lpa) soybean mutants. *Journal of Agricultural and Food Chemistry* 57: 6408-6416.
- Gaudreault, P-R, Webb JA (1981) Stachyose synthesis in leaves of *Cucurbita pepo*. *Phytochemistry* 20: 2629-2633.

- Gitzelmann R, Auricchio S (1965) The handling of soya alpha-galactosides by a normal and a galactosemic child. *Pediatrics* 36: 231-235.
- Goldflus F (2001) Ingredientes derivados do processamento da soja aplicados a nutrição animal. In: Simpósio sobre Ingredientes na Alimentação Animal, 1. Campinas, SP. Anais. Campinas: CBNA, p.1-70.
- Guildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nature* 10: 94-108.
- Guimarães VM, De Rezende ST, Moreira MA, Barros EG, Felix CR (2001) Characterization of α -galactosidases from germinating soybean seed and their use for hydrolysis of oligosaccharides. *Phytochemistry* 58: 67-73.
- Guo S, Kemphues KJ (1995) Par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 81: 611-620.
- Haagenson DM, Klotz KL, Campbell L (2008) Impact of storage temperature, storage duration, and harvest date on sugarbeet raffinose metabolism. *Postharvest Biology and Technology* 49: 221–228.
- Hannah MA, Zuther E, Buchel K, Heyer AG (2006) Transport and metabolism of raffinose family oligosaccharides in transgenic potato. *Journal of Experimental Botany* 57: 3801–3811.
- Hannon GJ (2002) RNA interference. *Nature* 418: 244-251.
- Haritatos E, Keller F, Turgeon R (1996) Raffinose oligosaccharide concentrations measured in individual cell and tissue types in *Cucumis melo* L. leaves: implications for phloem loading. *Planta* 198: 614–622.
- Hartwig EE, Kuo TM, Kenty MM (1997) Seed protein and its relationship to soluble sugars in soybean. *Crop Science* 37: 770–773.
- Hirai S, Kodama, H (2008) RNAi vectors for manipulation of gene expression in higher plants. *The Open Plant Science Journal* 2: 21-30.
- Hitz WD, Carlson TJ, Kerr PS, Sebastian SA (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiology* 128: 650-660.
- Hitz WD, Sebastian SA, Grace DJ, Streit LG (2004) Soybean plant producing seeds with reduced levels of raffinose saccharides and phytic acid. United States, Patente 20030074685. Data da publicação: 07 Janeiro.

- Hoch G, Peterbauer T, Richter, A. (1999). Purification and characterization of stachyose synthase from lentil (*Lens culinaris*) seeds: Galactopinitol and Stachyose Synthesis. *Archives of Biochemistry and Biophysics* 366: 75-81.
- Holthaus U, Schmitz K (1991a) Stachyose synthesis in mature leaves of *Cucumis melo*. Purification and characterization of stachyose synthase (EC 2.4.1.67). *Planta* 184: 525-531.
- Holthaus U, Schmitz K (1991b) Distribution and immunolocalization of stachyose synthase in *Cucumis melo* L. *Planta* 185: 479-486.
- Horbowicz M, Obendorf RL (1994) Seed desiccation tolerance and storability: Dependence on flatulence producing oligosaccharides and cyclitols – review and survey. *Seed Science Research* 4: 385–405.
- Hymowitz T, Walker WM, Collins FI, Panczner J (1972a) Stability of sugar content in soybean strains. *Communications in Soil Science and Plant Analysis* 3: 367–373.
- Hymowitz T, Walker WM, Collins FI, Panczner J (1972b) Relationship between the content of oil, protein, and sugar in soybean seed. *Agronomy Journal* 64: 613–615.
- Ibrahim HMM, Alkharouf NW, Meyer SLF, Aly MAM, El-Din AEFYG, Hussein EHA, Matthews BF (2010) Post-transcriptional gene silencing of root-knot nematode in transformed soybean roots. *Experimental Parasitology* 127: 90-99.
- Inan HC, Keller F (2002) Purification and characterization of the raffinose oligosaccharide chain elongation enzyme, galactan: galactan galactosyltransferase (GGT), from *Ajuga reptans* leaves. *Seed Science Research* 9: 305-310.
- Islam SMS, Miyazaki T, Tanno T, Itoh T (2005) Dissection of gene function by RNA silencing. *Plant Biotechnology* 22: 443–446.
- Jones DA, Dupont MS, Ambrose MJ, Frias J, Hedley LC (1999) The discovery of compositional variation for the raffinose family of oligosaccharides in pea seeds. *Seed Science Research* 9: 305-310.
- Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology* 57: 19-53.
- Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. *Plant Physiology* 136: 4159–4168.
- Karlsson H, Carlstedt I, Hansson GC (1989) The use of gas chromatography and gas chromatography-mass spectrometry for the characterization of permethylated

- oligosaccharides with molecular mass up to 2300. *Analytical Biochemistry* 182: 438-446.
- Karner U, Peterbauer T, Raboy V, Jones DA, Hedley CL, Richter A (2004) A myo-inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. *Journal of Experimental Botany* 55: 1981-1987.
- Karr-Lilienthal LK, Kadzere CT, Grieshop CM, Fahey JR., GC (2005) Chemical and nutritional properties of soybean carbohydrates as related to nonruminant. A review: *Livestock Production Science* 97: 1-12.
- Kavi HH, Fernandez H, Xie W, Birchler JA (2008) Genetics and biochemistry of RNAi in *Drosophila*. *Current Topics in Microbiology and Immunology* 320: 37-75.
- Kennerdell JR, Carthew RW (2000) Heritable gene silencing in *Drosophila* using double stranded RNA. *Nature Biotechnology* 18: 896–898.
- Kerr PS, Pearlstein RW, Schweiger BJ, Beckermanley MF, Pierce JW (1993). Nucleotide sequences of galactinol synthase from zucchini and soybean. International Patent Publication WO 93/02196, PCT/US92/06057.
- Ketting RF, Haverkamp TH, Van Luenen HG, Plasterk RH (1999) Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99: 133-141.
- Kolida S, Tuohy K, Gibson GR (2002) Prebiotic effects of inulin and oligofructose. *British Journal of Nutrition* 87: 193–197.
- Koster KL, Leopold AC (1988) Sugars and desiccation tolerance in seeds. *Plant Physiology* 88: 829-832.
- Kotani H, Nakamura Y, Sato S, Kaneko T, Asamizu E, Miyajima N, Tabata S (1997) Structural analysis of *Arabidopsis thaliana* chromosome 5. II. Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned P1 clones. *DNA Research* 31: 291-300.
- Kotiguda G, Peterbauer T, Mulimani VH (2006) Isolation and structural analysis of ajugose from *Vigna mungo* L. *Carbohydrate Research* 341: 2156–2160.
- Krishnan HB (2005) Engineering soybean for enhanced sulfur amino acid content. *Crop Science* 45: 454-461.
- Ku S, Wei LS, Steimberg MP, Nelson AI, Hymowitz T (1976) Extraction of oligosaccharides during cooking of whole soybean. *Journal of Food Science* 41: 361-364.

- Kumar V, Rani A, Goyal L, Dixit AK, Manjaya JG, Dev J, Swamy M (2010) Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *Journal of Agricultural and Food Chemistry* 58: 5081-5085.
- Kusaba M, Miyahara K, Iida S, Fukuoka H, Takano T, Sassa H, Nishimura M, Nishio T (2003) Low glutelin content1: a dominant mutation that suppresses the glutelin multigene family via RNA silencing in rice. *Plant Cell* 15: 1455–1467.
- L'Hocine L, Boye JI (2007) Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Critical Reviews in Food Science and Nutrition* 47: 127–143.
- Lahuta LB, Górecki RJ (2010) Raffinose in seedling of winter vetch (*Vicia villosa* Roth.) under osmotic stress and followed by recovery. *Acta Physiologiae Plantarum* DOI 10.1007/s11738-010-0597-4 (Article In Press).
- Lahuta LB, Górecki RJ, Michalczyk D, Piotrowicz-Cieslak AI (2000) Alpha-D-galactosidase activity in stored yellow lupin (*Lupinus luteus* L.) seeds. *Electronic Journal of Polish Agricultural Universities* 3: 1-10.
- Lehle L, Tanner W (1973) The function of myo-inositol in the biosynthesis of raffinose. *European Journal of Biochemistry* 38: 103– 110.
- Leske KL, Akavanichan O, Cheng TK, Coon CN (1991) Effect of ethanol extraction on nitrogen-corrected true metabolizable energy for soybean meal with broilers and roosters. *Poultry Science* 70: 892-895.
- Li M, Rohrer B (2006) Gene silencing in *Xenopus laevis* by DNA vector based RNA interference and transgenesis. *Cell Research* 16: 99–105.
- Li S, Li T, Kim W-D, Kitaoka M, Yoshida S, Nakajima M, Kobayashi H (2007) Characterization of raffinose synthase from rice (*Oryza sativa* L. var. Nipponbare). *Biotechnology Letters* 29: 635–640.
- Lichtenstein AH (1998) Soy protein, isoflavones and cardiovascular disease risk. *Journal of Nutrition* 128: 1589-1592.
- Liener IE (1994) Implications of antinutritional components in soybean foods. *Critical Reviews in Food Science and Nutrition* 34: 31–67.
- Lin TP, Huang NH (1994) The relationship between carbohydrate composition of some tree seeds and their longevity. *Journal of Experimental Botany* 45: 1289–1294.
- Lin X, Kaul S, Rounsley SD, Shea TP, Benito M-I, Town CD, Fujii CY, Mason TM, Bowman CL, Barnstead ME, Feldblyum TV, Buell CR, Ketchum KA, Lee JJ,

- Ronning CM, Koo HL, Moffat KS, Cronin LA, Shen M, Pai G, Van Aken S, Umayam L, Tallon LJ, Gill JE, Adams MD, Carrera AJ, Creasy TH, Goodman HM, Somerville CR, Copenhaver GP, Preuss D, Nierman WC, White O, Eisen JA, Salzberg SL, Fraser CM, Venter JC (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761-768.
- Liu H-L, Dai X-L, Xu Y-Y, Chong K (2007) Over-expression of OsUGE-1 altered raffinose level and tolerance to abiotic stress but not morphology in *Arabidopsis*. *Journal of Plant Physiology* 164: 1384-1390.
- Liu J-J, Krenz DC, Glavez AF, De Lumen BO (1998) Galactinol synthase (GS): increased enzyme activity and levels of mRNA due to cold and desiccation. *Plant Science* 134: 11–20.
- Liu J-J, Odegard W, De Lumen BO (1995) Galactinol synthase from kidney bean cotyledon and zucchini leaf. *Plant Physiology* 109: 505–511.
- Liu Q, Singh SP, Green AG (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiology* 129: 1732-1743.
- López-Gomollón S, Dalmay T (2011) Recent Patents in RNA Silencing in Plants: Constructs, Methods and Applications in Plant Biotechnology. *Recent patents on DNA & gene sequences* (Article In Press).
- Mahmood-Ur-Rahman, Ali I, Husnain T, Riazuddin S (2008) RNA interference: The story of gene silencing in plants and humans. *Biotechnology Advances* 26: 202-209.
- Mansoor S, Amin I, Hussain M, Zafar Y, Briddon RW (2006) Engineering novel traits in plants through RNA interference. *Trends in Plant Science* 11: 559-565.
- Martínez-Villaluenga C, Frías J, Vidal-Valverde C (2006) Funcional Lupin seeds (*Lupinus albus* L. and *Lupinus luteus* L.) after extraction of α -galactosides. *Food Chemistry* 98: 291-299.
- Matzke M, Matzke AJM, Kooter JM (2001) RNA: Guiding Gene Silencing. *Science* 293: 1080-1083.
- Mccaskill A, Turgeon R (2007) Phloem loading in *Verbascum phoeniceum* L. depends on the synthesis of raffinose-family oligosaccharides. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19619–19624.
- Mendoza C, Viteri FE, Lonnerdal BL, Young KA, Raboy V, Brown KH (1998) Effect of genetically modified, low-phytic acid maize on absorption of iron from tortillas. *American Journal of Clinical Nutrition* 68: 1123–1128.

- Messina M, Messina V (1991) Increasing use of soyfoods and their potential role in cancer prevention. *Journal of the American Dietetic Association* 91: 836-840.
- Minorsky P (2003) The hot and the classic. *Plant Physiology* 131: 1159-1160.
- Morais AAC, Silva AL (2000) Valor nutritivo e funcional da soja. *Revista Brasileira de Nutrição Clínica* 15: 306-315.
- Moreira MA (1999) Programa de melhoramento genético da qualidade de óleo e proteína de soja desenvolvido na UFV. In: Congresso Brasileiro de Soja, Londrina, PR. Anais, EMBRAPA-SPI, p. 99-104.
- Mueller S, Gausson V, Vodovar N, Deddouche S, Troxler L, Perot J, Pfeffer S, Hoffmann JA, Saleh MC, Imler JL (2010) RNAi-mediated immunity provides strong protection against the negative-strand RNA vesicular stomatitis virus in *Drosophila*. *Proceedings of the National Academy of Sciences* 107: 19390-19395.
- Neff WE, Selke E, Mounts TL, Rinsch W, Frankel EN, Zeitoun MAM (1992) Effect of triacylglycerol composition and structures on oxidative stability of oils from selected soybean germplasm. *Journal of the American Oil Chemists' Society* 69: 111-118.
- Neus JD, Fehr WF, Schnebly SR (2005) Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. *Crop Science* 45: 589-592.
- Neus JD, Fehr WF, Schnebly SR (2005) Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. *Crop Science* 45: 589-592.
- Nishizawa A, Yabuta Y, Shigeoka S (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiology* 147: 1251-1263.
- Nunes ACS, Vianna GR, Cuneo F, Amaya-Farfán J, Capdeville G, Rech EL, Aragão FJL (2006) RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (*GmMIPS1*) in transgenic soybean inhibited seed development and reduced phytate content. *Planta* 224: 125-132.
- Obendorf RL (1997) Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Science Research* 7: 63-74.
- Obendorf RL, Horbowicz M, Dickerman AM, Brenac P, Smith ME (1998) Soluble oligosaccharides and galactosyl cyclitols in maturing soybean seeds in plant and in vitro. *Crop Science* 38: 78-84.

- Obendorf RL, Odorcic S, Ueda T, Coseo MP, Vassallo E (2004) Soybean galactinol synthase forms fagopyritol B1 but not galactopinitols: substrate feeding of isolated embryos and heterologous expression. *Seed Science Research* 14: 321–333.
- Obendorf RL, Sensenig EM, Wu J, Ohashi M, O'Sullivan, TE, Kosina SM, Schnebly SR (2008) Soluble carbohydrates in mature soybean seed after feeding D-chiro-inositol, myo-inositol, or D-pinitol to stem-leaf-pod explants of low-raffinose, low-stachyose lines. *Plant Science* 175: 650-655.
- Ogita S, Uefuji H, Morimoto M, Sano H (2004) Application of RNAi to confirm theobromine as the major intermediate for efficient biosyntheses in coffee plants with potential for construction of decaffeinated varieties. *Plant Molecular Biology* 54: 931-941.
- Osaki M, Batalha MO (2008) Produção de biodiesel e óleo vegetal no Brasil: Realidade e Desafio, In: SOBER – XLVI Congresso da Sociedade Brasileira de Economia, Administração e Sociologia rural. São Carlos, São Paulo.
- Panikulangara TJ, Eggers-Schumacher G, Wunderlich M, Stransky H, Schöffl F (2004) Galactinol synthase1: a novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in *Arabidopsis*. *Plant Physiology* 136: 3148–3158.
- Parrott W, Chassy B, Ligon J, Meyer L, Petrick J, Zhou J, Herman R, Delaney B, Levine M (2010) Application of food and feed safety assessment principles to evaluate transgenic approaches to gene modulation in crops. *Food and Chemical Toxicology* 48: 1772-1790.
- Parsons CM, Zhang Y, Araba M (2000) Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poultry Science* 79: 1127–1131.
- Pennycooke JC, Jones ML, Stushnoff C (2003) Down-regulating α -galactosidase enhances freezing tolerance in transgenic petunia. *Plant Physiology* 133: 901–909.
- Perrimon N, Ni JQ, Perkis L (2010) In vivo RNAi: today and tomorrow. *Cold Spring Harbor Perspectives in Biology* 2: a003640.
- Peterbauer T, Karner U, Mucha J, Mach L, Jones DA, Hedley CL, Richter A (2003) Enzymatic control of the accumulation of verbascose in pea seeds. *Plant Cell and Environment* 26: 1385-1391.
- Peterbauer T, Mach L, Mucha J, Richter A (2002a) Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the

- enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. *Planta* 215: 839-846.
- Peterbauer T, Mucha J, Mach L, Richter A (2002b) Chain-elongation of raffinose in pea seeds. Isolation, characterization, and molecular cloning of multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. *Journal of Biological Chemistry* 277: 194–200.
- Peterbauer T, Mucha J, Mayer U, Popp M, Glossl J, Richter A (1999) Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*): molecular cloning and functional expression of stachyose synthase. *The Plant Journal* 20: 509-518.
- Peterbauer T, Richter A (1998) Galactosylononitol and stachyose synthesis in seeds of adzuki bean. Purification and characterization of stachyose synthase. *Plant Physiology* 117: 165-172.
- Peterbauer T, Richter A (2001) Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research* 11: 185-197.
- Peters S, Mundree SG, Thomson JA, Farrant JM, Keller F (2007) Protection mechanisms in the resurrection plant *Xerophyta viscosa* (Baker): both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit. *Journal of Experimental Botany* 58: 1947–1956.
- Philippe RN, Ralph SG, Mansfield SD, Bohlmann J (2010) Transcriptome profiles of hybrid poplar (*Populus trichocarpa* × *deltoides*) reveal rapid changes in undamaged, systemic sink leaves after simulated feeding by forest tent caterpillar (*Malacosoma disstria*). *New Phytologist* 188: 787-802.
- Pooggin M, Shivaprasad PV, Veluthambi K, Hohn T (2003) RNAi targeting of DNA virus in plants. *Nature Biotechnology* 21:131–132.
- Pridham JB, Hassid WZ (1965) Biosynthesis of raffinose. *Plant Physiology* 40: 984–986.
- Pugalenthi M, Siddhuraju V, Vadil V (2006) Effect of soaking followed by cooking and the addition of α -galactosidase on oligosaccharides levels in different *Canavalia* accessions. *Journal of Food Composition and Analysis* 19: 512-517.
- Qi Y, Denli AM, Hannon GJ (2005) Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Molecular Cell* 19: 421–428.

- Raboy V (1997) Accumulation and storage of phosphate and minerals. In: Larkins, B.A., Vasil, I.K. (Eds.), Cellular and Molecular Biology of Plant Seed Development. Kluwer Academic, Dordrecht, Netherlands, pp. 441–477.
- Rackis JJ (1981) Flatulence caused by soy and its control through processing. Journal of the American Oil Chemists' Society. 58: 503-505.
- Rehms H, Barz W (1995) Degradation of stachyose, raffinose, melibiose and sucrose by different tempe-producing *Rhizopus fungi*. Applied Microbiology and Biotechnology 44: 47-52.
- Ribeiro M, Felix CR, LOZZI, S de P (2000) Soybean seed galactinol synthase activity as determined by a novel colorimetric assay. Revista Brasileira de Fisiologia Vegetal, 12: 203-212.
- Richter A, Hoch G, Puschenreiter M, Mayer U, Peterbauer T (2000) The role of stachyose synthase in the oligosaccharide metabolism of legume seeds. pp. 75–84 in Black, M., Bradford, K.J., Vásquez-Ramos, J. (Eds) Seed biology. Advances and applications. Wallingford, UK, CAB International.
- Riedmann LT, Schwentner R (2010) miRNA, siRNA, piRNA and argonautes: news in small matters. RNA Biology 7: 133-139.
- Runo S, Alakonya A, Machuka J, Sinha N (2010) RNA interference as a resistance mechanism against crop parasites in Africa: a 'Trojan horse' approach. Pest Management Science 67: 129-136.
- Ruttloff H, Taufel A, Krause W, Haenel H, Taufel K (1967) The intestinal enzymatic decomposition of galacto-oligosaccharides in the human and animal intestine, with particular regard to *Lactobacillus bifidus*. Part II. On the intestinal behaviour of lactulose. Die Nahrung 11: 39-46.
- Sanni AI, Onilude AA, Ogundoye OR (1997) Effect of bacterial galactosidase treatment on the nutritional status of soybean seed and its milk derivative. Alimentaria 33: 127-132.
- Santos MAT (2006) Effect of boiling on contents of antinutritional factors in leaves of broccoli, cauliflower and cabbage. Ciência e Agrotecnologia 30: 294-301.
- Santos TB, Budzinski IGF, Marur CJ, Petkowicz CLO, Pereira LFP, Vieira LGE (2011) Expression of three galactinol synthase isoforms in *Coffea arabica* L. and accumulation of raffinose and stachyose in response to abiotic stresses. Plant Physiology and Biochemistry, doi:10.1016/j.plaphy.2011.01.023 (Article In Press).

- Saravitz DM, Pharr DM, Carter TE (1987) Galactinol synthase activity and soluble sugars in developing seeds of four soybean genotypes. *Plant Physiology* 83: 185–189.
- Schillinger JA, Dierking EC, Bilyeu, KD (2011) Soybeans having high germination rates and ultra-low raffinose and stachyose content. United States Patent Application 20110003045, A1.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean. *Nature* 463: 178-183.
- Sebastian SA, Kerr PS, Pearlstein RW, Hitz WD (2000) Soybean germplasm with novel genes for improved digestibility. pp. 56–73 in Drackley, J.K. (Ed.) *Soy in animal nutrition*. Savoy, Illinois, Federation of Animal Science Societies.
- Sebastian SA, Kerr PS, Pearlstein RW, Hitz WD (2000) Soybean germplasm with novel genes for improved digestibility. In: Drackley, JK (Ed.). *Soy in animal nutrition*. Federation of Animal Science Societies, Savoy, IL, p.56-74.
- Segal G, Song R, Messing J (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. *Genetics* 165: 387-397.
- Senthil-Kumar M, Mysore KS (2011) Virus-induced gene silencing can persist for more than 2 years and also be transmitted to progeny seedlings in *Nicotiana benthamiana* and tomato. *Plant Biotechnology Journal*, doi: 10.1111/j.1467-7652.2011.00589.x. (Article In Press).
- Silva J, Chang K, Hannon GJ, Rivas FV (2004) RNA-interferencebased functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* 23: 8401–8409.
- Smiricky HR, Grieshop CH, Albin DM, Wubben JE, Gabert VM, Fahey GCJ (2002) The influence of soy oligosaccharides on apparent and true ideal amino acid digestibilities and fecal consistency in growing pigs. *Journal of Animal Science* 80: 2433-2441.
- Smith NA (2000) Total silencing by a intron-spliced hairpin RNA. *Nature* 407: 319-320.

- Smith PT, Kuo TM, Crawford CG (1991) Purification and characterization of galactinol synthase from mature zucchini squash leaves. *Plant Physiology* 96: 693-698
- Smits CHM, Annison G (1996) Non-starch plant polysaccharides in broiler nutrition towards a physiologically valid approach to their determination towards a physiologically valid approach to their determination. *World's Poultry Science Journal* 52: 204-221.
- Soosaar JL, Burch-Smith TM, Dinesh-Kumar SP (2005) Mechanisms of resistance to viruses. *Nature Reviews Microbiology* 3: 789-798.
- Sprenger N, Keller F (2000) Allocation of raffinose family oligosaccharides to transport and storage pools in *Ajuga reptans*: the roles of two distinct galactinol synthases. *The Plant Journal* 21: 249–258.
- Styer JC, Keddie J, Spence J, Gillaspay GE (2004) Genomic organization and regulation of the LeIMP-1 and LeIMP-2 genes encoding myo-inositol monophosphatase in tomato. *Gene* 326: 35–41.
- Suarez FL, Springfield J, Furne JK, Lohrmann TT, Kerr PS, Levit MD (1999) Gas production in humans ingesting soybean flour derived from beans naturally low in oligosaccharides. *American Journal of Clinical Nutrition* 69:135–139.
- Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic. *Proceedings of the National Academy of Sciences* 103: 18054-18059.
- Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC (1999) The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99: 123-132.
- Taji T, Ohsumi C, Luchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki M, Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. *The Plant Journal* 4: 417-426.
- Takanashi K, Shitan N, Sugiyama A, Kamimoto Y, Hamamoto M, Iwaki T, Takegawa K, Yazaki K (2008) Galactinol Synthase Gene of *Coptis japonica* Is Involved in Berberine Tolerance. *Bioscience, Biotechnology and Biochemistry* 72: 389-405.
- Tang G, Galili G (2004) Using RNAi to improve plant nutritional value: from mechanism to application. *Trends in Biotechnology* 22: 463-469.

- Tang G, Galili G, Zhuang X (2007) RNAi and microRNA: breakthrough Technologies for the improvement of plant nutritional value and metabolic engineering. *Metabolomics* 3: 357-369.
- Tanner W, Kandler O (1968) Myo-Inositol, a cofactor in the biosynthesis of stachyose. *European Journal of Biochemistry* 4: 233–239.
- Tavares GM, Pitanga KC, Silva MAT, Morais AAC (2000) A soja como alimento funcional na prevenção do câncer. *Revista Brasileira de Nutrição Clínica* 15: 321-325.
- Tenllado F, Martínez-García B, Vargas M, Díaz-Ruiz JR (2003) Crude extracts of bacterially expressed dsRNA can be used to protect plants against virus infections. *BMC Biotechnology* 3: 3–14.
- Trugo LC, Farah A, Cabral L (1995) Oligosaccharide distribution in Brazilian soy bean cultivars. *Food Chemistry* 52: 385-387.
- United States Department of Agriculture (USDA) - World Agricultural Supply and Demand Estimates - ISSN: 1554-9089. Disponível em: <http://www.usda.gov/oce/commodity/wasde/latest.pdf>. Acesso em: 12/02/2011.
- Vanitharani R, Chellappan P, Fauquet CM (2003) Short interfering RNA-mediated interference of gene expression and viral DNA accumulation in cultured plant cells. *Proceedings of National Academy of Sciences of USA* 100: 9632–9636.
- Vaucheret H, Beclin C, Fagard M (2001) Post-transcriptional gene silencing in plants. *Journal of Cell Science* 114: 3083-3091.
- Voitsekhovskaja OV, Rudashevskaya EL, Demchenko KN, Pakhomova MV, Batashev DR, Gamalei YV, Lohaus G, Pawlowski K (2009). Evidence for functional heterogeneity of sieve element–companion cell complexes in minor vein phloem of *Alonsoa meridionalis*. *Journal of Experimental Botany*. 60:1873-1883.
- Wagner JR, Becker R, Gumbmann MR, Olson AC (1976) Hidrogen production in the rat following ingestion of raffinose, stachyose, and oligosaccharide-free bean residue. *Journal Chemistry Technology* 56: 3-13.
- Wang G, Xu Y (2008) Hypocotyl-based *Agrobacterium*-mediated transformation of soybean (*Glycine max*) and application for RNA interference. *Plant Cell Reports* 27: 1177–1184.
- Wang MB, Waterhouse PM (2001) Application of gene silencing in plants. *Current Opinion in Plant Biology* 5: 146-150.

- Watanabe S, Nakagawa A, Izumi S, Shimada H, Sakamoto A (2010) RNA interference-mediated suppression of xanthine dehydrogenase reveals the role of purine metabolism in drought tolerance in *Arabidopsis*. *FEBS Letters* 584: 1181-1186.
- Waterhouse PM, Grabam MW, Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proceedings of National Academy of Sciences of USA* 95: 13959-13964.
- Watson JM, Furaso AF, Wang M, Waterhouse PM (2005) RNA silencing platforms in plants. *FEBS Letters* 579: 5982-5987.
- Webb JA (1982) Partial purification of galactinol synthase from leaves of *Cucurbita pepo*. *Canadian Journal of Botany* 60: 1054-1059.
- Wesley SV, Helliwell CA, Smith NA, Wang M, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *The Plant Journal* 27: 581-590.
- Wielopolska A, Townley H, Moore I, Waterhouse P, Helliwell C (2005) A high-throughput inducible RNAi vector for plants. *Plant Biotechnology Journal* 3: 583-590.
- Wiggins HS (1984) Nutritional value of sugar anal related compounds undigested in the small gut. *Proceeding of Nutrition Society* 43: 69-75.
- Willians M, Clark G, Sathasivan K, Islam AS (2004) RNA interference and its application in crop improvement. Disponível em: <http://www.sbs.utexas.edu/islam/RNAi-ReviewArticle%20-PTC-Final-ubmitted.pdf>. Acessado em: 15/02/2011.
- Wu X, Kishitani S, Ito Y, Toriyama K (2009) Accumulation of raffinose in rice seedlings overexpressing OsWRKY11 in relation to desiccation tolerance. *Plant Biotechnology* 26: 431-434.
- Xiong AS, Yao Q-H, Peng P-H, Xian L, Han P-L, Fan, H-Q (2005) Different effect on ACC oxidase gene silencing triggered by RNA interference in transgenic tomato. *Plant Cell Reports* 23: 639-646.
- Yadav BC, Vethambi K, Subramaniam K (2006) Host-generated double-stranded RNA induces RNAi in plant-parasitic nematodes and protects the host form infection. *Molecular and Biochemical Parasitology* 148: 219-222.

- Yadav NS (1996) Genetic modification of soybean oil quality. In Verma, DPS, Shoemaker, RC (Eds) Soybean genetics, molecular biology and biotechnology. USA: CAB INTERNATIONAL, p.127-188.
- Yu H, Kumar PP (2003) Post-transcriptional gene silencing in plants by RNA. Review, published online: 21 august ©Springer-Verlag.
- Yuan FJ, Zhu DH, Deng B, Fu XJ, Dong DK, Zhu SL, Li BQ, Shu QY (2009) Effects of two low phytic acid mutations on seed quality and nutritional traits in soybean (*Glycine max* L. Merr). Journal of Agricultural and Food Chemistry 57: 3632-3638.
- Zamore PD (2002) Ancient pathways programmed by small RNAs. Science 296: 1265-1269.
- Zamore PD, Tuschl TS, Sharp PA, Bartel DP (2000) RNAi: Double-stranded RNA directs the ATP dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101: 25-33.
- Zarkadas CG, Gagnon C, Gleddie S, Khanizadeh S, Cober ER, Guillemette RJD (2007) Assessment of the protein quality of fourteen soybean [*Glycine max* (L.) Merr.] cultivars using amino acid analysis and two-dimensional electrophoresis. Food Research International 40: 129-146.
- Zhao MM, An DR, Zhao J, Huang GH, He ZH, Chen JY (2006) Transiently expressed short hairpin RNA targeting 126 kDa protein of tobacco mosaic virus interferes with virus infection. Acta Biochimica et Biophysica Sinica 38: 22–28.
- Zhao T-Y, Thacker R, Corum JW, Snyder JC, Meeley RB, Obendorf RL, Downie B. (2004) Expression of the maize Galactinol Synthase gene family: (I) Expression of two different genes during seed development and germination. Physiologia Plantarum 121: 634-646.
- Zuther E, Bachel K, Hundertmark M, Stitt M, Hinch DK, Heyer AG (2004) The role of raffinose in the cold acclimation response of *Arabidopsis thaliana*. FEBS Letters 576: 169–173.

4. OBJETIVOS

4.1. Objetivos gerais

Os objetivos deste trabalho foram identificar e caracterizar *in silico* os genes que codificam enzimas envolvidas na biossíntese de oligossacarídeos de rafinose, além de construir cassetes de expressão visando o silenciamento do gene da estaquiose sintase, via RNA de interferência, em soja (*Glycine max* L. Merrill).

4.2. Objetivos específicos

- Identificar e caracterizar *in silico* os genes que codificam as enzimas envolvidas na biossíntese de ROs em soja;
- Analisar a expressão *in silico* dos genes que codificam as enzimas envolvidas na biossíntese desses oligossacarídeos;
- Construir cassetes para expressão constitutiva de hpRNA em soja;
Construir cassetes para expressão semente específica de hpRNA em soja.

CAPÍTULO 1

***IN SILICO* CHARACTERIZATION AND ANALYSIS OF GENE EXPRESSION INVOLVED IN THE BIOSYNTHESIS OF RAFFINOSE OLIGOSACCHARIDES IN SOYBEAN (*Glycine max* L. Merrill)**

ABSTRACT

Raffinose oligosaccharides (ROs) are synthesized by a group of enzymes whose functions are well understood. However, the genes and the expression patterns of these enzymes have not been fully described in soybean (*Glycine max* L. Merrill). This study aimed to characterize and analyze in silico the expression of genes responsible for ROs biosynthesis in soybean. BLAST analysis against the soybean genome (Glyma1) revealed 10 potential genes encoding ROs synthases. Based on phylogenetic analysis and the presence of conserved domains in the predicted proteins we detected six putative galactinol synthases, three raffinose synthases and one stachyose synthase. Additionally, we identified and characterized seven putative genes encoding alkaline α -galactosidases, also known as seed imbibition proteins. Most of the predicted proteins contain no signal peptide and are likely to be cytoplasmic. The exon-intron structure of genes was determined based on the genome. The sequences were positioned on the physical map of soybean and their promoter regions were analyzed. Cis-acting elements responsible for seed-specific expression and responses to biotic and abiotic stresses were identified. The virtual expression analysis of the putative genes for galactinol synthases and raffinose synthases indicate that these genes are mainly expressed in seeds, leaves and stressed roots. These analyses also suggest that the expression level of the potential stachyose synthase gene is higher in leaves of stressed plants than in seeds

and that the putative genes for alkaline α -galactosidases are abundantly expressed in seeds.

1. INTRODUCTION

Raffinose oligosaccharides (ROs) are a group of carbohydrates that accumulate in leaves during photosynthesis (Senser and Kandler 1967), in seeds during maturation (Lowell and Kuo 1989) and which serve as transport carbohydrates in the phloem in some plant families (Voitsekhovskaja et al. 2009). In leaves, ROs have been recognized as protective agents against abiotic stress-induced damage (Taji et al. 2002; Nishizawa et al. 2008). ROs have been correlated with plant vigor, viability and storability (Haagenson et al. 2008), as well as with desiccation tolerance during seed maturation (Obendorf 1997).

ROs are accumulated in significant amounts in soybean seed (Schweizer et al. 1978), though their roles are not well elucidated in this legume. Identification and expression pattern analysis of genes involved in ROs biosynthetic pathways may help understand the roles of these compounds in the mechanisms of response to environmental stress and seed desiccation tolerance in soybean.

The biosynthetic pathway of ROs is well characterized (Dey, 1985). It requires sucrose and occurs by the transfer of galactosyl groups originating from an appropriate donor, such as UDP-D-galactose, galactinol or ROs themselves, to an acceptor. The acceptor molecule can be sucrose, raffinose, stachyose or verbascose. The first step of the pathway is the formation of galactinol by the action of enzyme galactinol synthase (GolS; UDP-galactose:*myo*-inositol 3- α -D-galactosyltransferase; EC 2.4.1.123) from UDP-galactose and *myo*-inositol (Saravitz et al. 1987). Raffinose and stachyose are then synthesized by sequential addition of D-galactose units from galactinol to sucrose and raffinose, respectively. These reversible reactions are catalyzed by raffinose synthase (RFS; α -D-galactosyl-(1 \rightarrow 3)-1D-*myo*-inositol:sucrose 6- α -D-galactosyltransferase; EC 2.4.1.82) (Lehle and Tanner, 1973) and stachyose synthase (STS; α -D-galactosyl-

(1→3)-1D-*myo*-inositol:raffinose galactosyltransferase; EC 2.4.1.67) (Peterbauer and Richter 1998).

The objectives of this study were to identify, characterize and analyze the *in silico* expression of genes involved in the biosynthesis of ROs which respectively encode the enzymes GolS, RFS and STS in soybean.

2. MATERIALS AND METHODS

The identification of putative genes that encode for GolS, RFS and STS was performed using the TBLASTN algorithm, with a cutoff point E-value $< 1e^{-5}$, against the proteome predicted from the annotation of soybean genome database used seventeen described and/or characterized protein sequences (queries) belonging to nine plant species: *Arabidopsis thaliana* (Genbank: NP_182240 – AtGolS1; NP_176053 – AtGolS2; NP_172406 – AtGolS3; NP_176250 – AtGolS4; NP_197768 – AtGolS5; NP_567741 – AtGolS6; NP_176248 – AtGolS7), *Coffea arabica* (Genbank: ADM92588 – CaGolS1; ADM92590 – CaGolS2; ADM92589 – CaGolS3), *Glycine max* (Genbank: AAM96867 – GmGolS), *Populus trichocarpa* (Genbank: ACA04027 – PtGolS1 ; ACA04030 – PtGolS2; ACA04032 – PtGolS3) and *Zea mays* (Genbank: AAQ07248 – ZmGolS1; AAQ07249 – ZmGolS2; ACG39512 – ZmGolS3). The putative genes *GmRFS* were identified from ten the polypeptide sequences belonging to species: *Arabidopsis* (Genbank: BAB11595 – AtRFS), *Cucumis sativus* (Genbank: AAD02832 – CsRFS), *G. max* (Genbank: E25448– GmRFS1; ACD13461 – GmRFS2; ACD13462 – GmRFS3), *Oryza sativa* (Genbank: Q5VQG4 – OsRFS), *Pisum sativum* (Genbank: CAD20127 – PsRFS), *P. trichocarpa* (Genbank: XP_002321625 – PtRFS), *Vicia faba* (Genbank: E24423 – VfRFS) and *Vitis vinifera* (Genbank: XP_002281483 – VvRFS). Amino acid sequences of *GmRFS* and *VfRFS* were deduced from predicted coding sequence (CDS) using ORF Finder program (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The identification of the putative *GmSTS* genes was performed based on protein sequences of seven species: *Alonsoa meridionalis* (Genbank: CAD31704 – AmSTS), *Arabidopsis* (Genbank: NP_192106 – AtSTS), *C. sativus* (Genbank: ABQ53598 – CsSTS), *P. sativum* (Genbank: CAC38094 – PsSTS1; CAD55555 – PsSTS2), *Stachys affinis* (Genbank: CAC86963 – SaSTS),

Vigna angularis (Genbank: CAB64363 – VaSTS). The sequences of soybean genes identified in this study were obtained using Phytozome GBrowser. Several putative genes located on different chromosomes were found. Redundant and incomplete genes were removed manually. The alternative transcripts of each gene were analyzed and those who not had at least one of each conserved domain superfamily in its sequence were removed. Predicted proteins were aligned using the ClustalW2 program (Thompson et al. 1994). Conserved domains in protein sequences corresponding to genes of interest were identified manually and using InterProscan, InterPro DB (<http://www.ebi.ac.uk/interpro/>) and SMART database (<http://smart.embl-heidelberg.de/>) (Letunic et al. 2004). Classification of the glycosyltransferases (GT) superfamily and glycosyl hydrolases (GH) superfamily was based on the Carbohydrate-Active EnZYme database (CAZy; <http://www.cazy.org/>). Predicted proteins that did not have at least one domino conserved characteristic of their superfamily were removed.

To build the physical map and to determine the exon-intron organization of genes, the consensus sequence of the soybean genome database available in SoyBase (<http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/#search>) (Grant et al. 2010) and the software GSDS (<http://gsds.cbi.pku.edu.cn/>) were used, respectively. On line software ProtParam (<http://expasy.org/tools/protparam.html>) was used to predict the molecular weight, PI and number of amino acids. A search for signal peptides was performed using the softwares SignalP (<http://www.cbs.dtu.dk/services/SignalP/>), Signal-3L (<http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/>) and SMART. The prediction of subcellular localization of putative proteins was performed using the software PREDOTAR (<http://urgi.versailles.inra.fr/predotar/predotar.html>) and WolfPsort (<http://wolfpsort.org/>). Multiple sequence alignments were carried out and a phylogenetic tree was constructed using MEGA version 4 (Tamura et al. 2007) applying the Neighbour-Joining (NJ) method (Saitou and Nei 1987) with bootstrap test of 1000 repetitions.

The identification of the presence of cis-regulatory elements within 2.0 kb regions upstream of the predicted coding sequence (CDS) was performed using database associated search tool in PLACE (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>) (Higo et al. 1999) and the software SignalScan (Prestridge 1991). The *in silico* analysis of expression of the genes involved in biosynthesis of raffinose oligosaccharides was carried out from the predicted CDS. These sequences were used to search, through *blastn* algorithms, the

NCBI expressed sequence tags (ESTs) soybean database composed of 1,461,623 sequences. Relative expression of each gene was associated with the number of hits obtained, considering the E-value $<1e^{-100}$ as cutoff. Redundant ESTs for each gene were removed.

3. RESULTS AND DISCUSSION

3.1. *In silico* identification and sequence analysis

A blast search of the soybean genome originally revealed the existence of seventeen putative genes encoding enzymes that could be involved in ROs biosynthesis. For standardization, these putative genes identified in the soybean genome, and distributed among twelve chromosomes (Fig. 1), were designated as: *GmGolS03.2*, *GmGolS03.3*, *GmGolS10*, *GmGolS19.2*, *GmGolS19.3*, *GmGolS20*, *GmRFS05.1*, *GmRFS05.2*, *GmRFS06*, *GmSTS19.1*, *GmSIP02*, *GmSIP03.1*, *GmSIP04*, *GmSIP09*, *GmSIP13*, *GmSIP14*, *GmSIP17* (Table 1). Examination of the location of each identified gene (<http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/>) revealed that all these genes, except *GmSIP02*, *GmRFS05.1* and *GmSIP14*, were located in duplicated regions (Table 1).

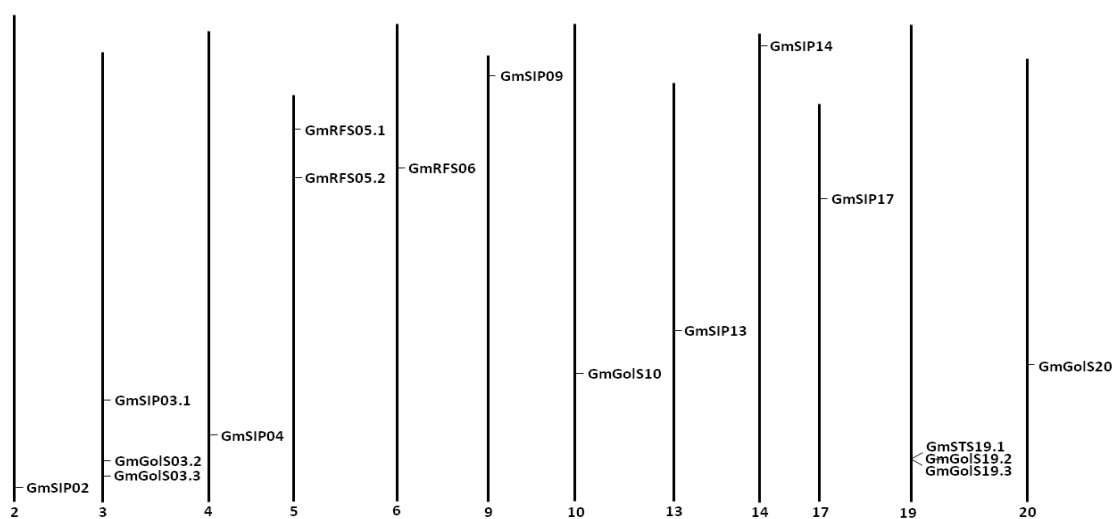


Fig. 1 Genomic distribution of the identified putative genes involved in ROs metabolism on *Glycine max*. The bars represent soybean chromosomes with the putative gene loci identified on the right (scale in Mbp)

Table 1 Summary of putative genes encoding enzymes involved in ROs metabolism

Gene Name	Locus	Protein Name	Protein			Duplicate Region ^a
			MM	AA	PI	
<i>GmGols03.2</i>	Glyma03g38080	GmGols03.2	38817.5	339	5.47	19
<i>GmGols03.3</i>	Glyma03g38910	GmGols03.3	38097.3	331	6.08	19
<i>GmGols10</i>	Glyma10g28610	GmGols10	38027.6	328	5.79	20
<i>GmGols19.2</i>	Glyma19g40680	GmGols19.2	38362.1	335	5.45	03
<i>GmGols19.3</i>	Glyma19g41550	GmGols19.3	38058.5	330	5.48	03
<i>GmGols20</i>	Glyma20g22700	GmGols20	37567.0	324	5.79	10
<i>GmSIP02</i>	Glyma02g47330	GmSIP02	83595.0	756	5.87	*
<i>GmSIP03.1</i>	Glyma03g29440	GmSIP03.1	82359.9	750	6.18	19
<i>GmSIP04</i>	Glyma04g36410	GmSIP04	84481.9	760	5.14	06
<i>GmSIP09</i>	Glyma09g01940	GmSIP09	79800.6	720	4.86	15
<i>GmSIP13</i>	Glyma13g22890	GmSIP13	82200.5	749	5.92	17
<i>GmSIP14</i>	Glyma14g01430	GmSIP14	83189.4	755	5.73	*
<i>GmSIP17</i>	Glyma17g11970	GmSIP17	82229.4	747	5.82	13
<i>GmRFS05.1</i>	Glyma05g02510	GmRFS05.1	85996.4	772	6.04	*
<i>GmRFS05.2</i>	Glyma05g08950	GmRFS05.2	81683.1	738	5.78	19
<i>GmRFS06</i>	Glyma06g18890	GmRFS06	85724.9	771	6.46	04
<i>GmSTS19.1</i>	Glyma19g40550	GmSTS19.1	95244.7	860	5.67	03

^a Duplicated region in Chromosomes

*Genes located outside of duplicated regions

Six identified genes in the soybean genome were classified, using SMART and InterProscan database, as glycosyltransferases (GT) family 8 (GT8): *GmGols03.2*, *GmGols03.3*, *GmGols10*, *GmGols19.2*, *GmGols19.3* and *GmGols20*; and were collectively referred to as *GmGols* genes. These genes have been described as a small gene family in different plant species, often expressed in response to abiotic stress (Taji et al. 2002; Zhao et al. 2003; Downie et al. 2003; Nishizawa et al. 2008; Santos et al. 2011).

Alignment of the amino acid sequences identified the presence of DxD and HxxGxxKPW motifs conserved across all identified predicted GT8 proteins (Fig. 2). These motifs have been suggested to be part of the catalytic site of all plant members of

the GT8 enzyme family (Yin et al. 2010). Furthermore, analysis of the deduced proteins, by using blastp against the non-redundant (NR) Genbank protein database revealed well-matched GolS sequences from a variety of plant sources. The deduced amino acid sequences of all identified GmGolS in the soybean genome share common features with those of other plant GolS: they have a molecular mass of about 38 kDa, they present GT family 8 domains, which is characteristic of GolS, and also the APSAA pentapeptide at the carboxy-terminal region (Fig. 2) (Taji et al. 2002; Downie et al. 2003; Takanashi et al. 2008; Santos et al. 2011). All GmGolS proteins identified in this work present a characteristic DxD/DGD motif, a binding site for manganese (Wang et al. 2009), important co-factor for their activity (Liu et al. 1995) (Fig. 2). Number of amino acid residues, molecular weight and pI, for each putative GmGolS protein, are shown in Table 1. Analysis of the subcellular localization of GmGolS proteins suggests that all these proteins are localized in the cytoplasm and that the mature proteins are devoid of N-terminal signal peptides, consistent with published reports (Bachmann and Keller 1995; Sprenger and Keller, 2000; Downie et al. 2003).

Based on the predicted CDS, we constructed a phylogenetic tree to analyze the relationship between GmGolS and other plant GT8 sequences deposited in the Genbank and the Phytozome databases (Fig. 3). Three main groups could be distinguished: family 8 of cell wall biosynthesis-related proteins (group I), including α -galacturonosyltransferase (GAUT) and GAUT-like proteins (GATL); glycogenin-like starch initiation proteins (PGSIPS) (group II), with PGSIPA-A and PGSIP-B; and GolS proteins (group III). The eight GmGolS proteins identified in this work clearly group with the other plant GolS proteins. Branch distribution shows that both, GmGolS03.2 and GmGolS19.2, formed a clade with rape (*Brassica napus*) (*BnGolS4*) and corn (*Z. mays*) (*ZmGolS1*, *ZmGolS2* and *ZmGolS3*). It also should be noted that GmGolS03.2 and GmGolS19.2 share 95% amino acid identity among themselves and about 75% with *BnGolS*. The other six GmGolS identified independently clustered in the same group. The deduced amino acid sequence of GmGolS19.3 were 94% identical to the sequence GmGolS03.3. Sequence GmGolS10 showed 93% amino acid identity with GmGolS20.

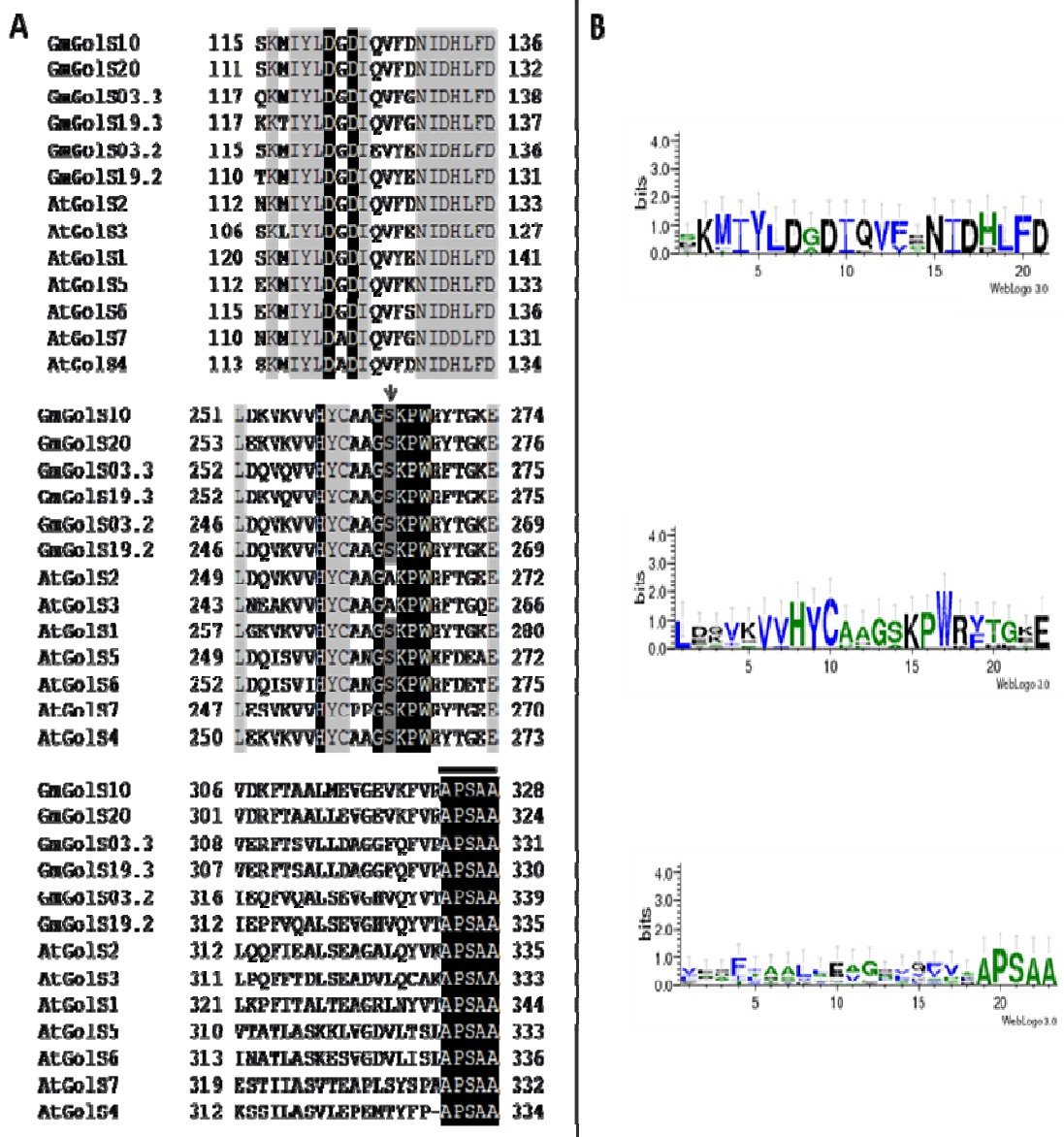


Fig. 2 Analysis of deduced GmGolS sequences encoded by the soybean genome. (A) Alignment of six *Glycine max* GolS (GmGolS03.2; GmGolS03.3; GmGolS10; GmGolS19.2; GmGolS19.3; GmGolS20) and seven GolS from *Arabidopsis* (NP_182240; NP_176053; NP_172406; NP_176250; NP_197768; NP_567741; NP_176248). Conserved motifs are highlighted in black and bold; a putative serine phosphorylation site is shown in dark gray by an arrow; and the characteristic hydrophobic pentapeptide (APSAA) is indicated by a bar; (B) Weblog output of the amino acid conservation of GolS enzymes aligned in the program WebLogo 3.0 (<http://weblogo.threeplusone.com/>)

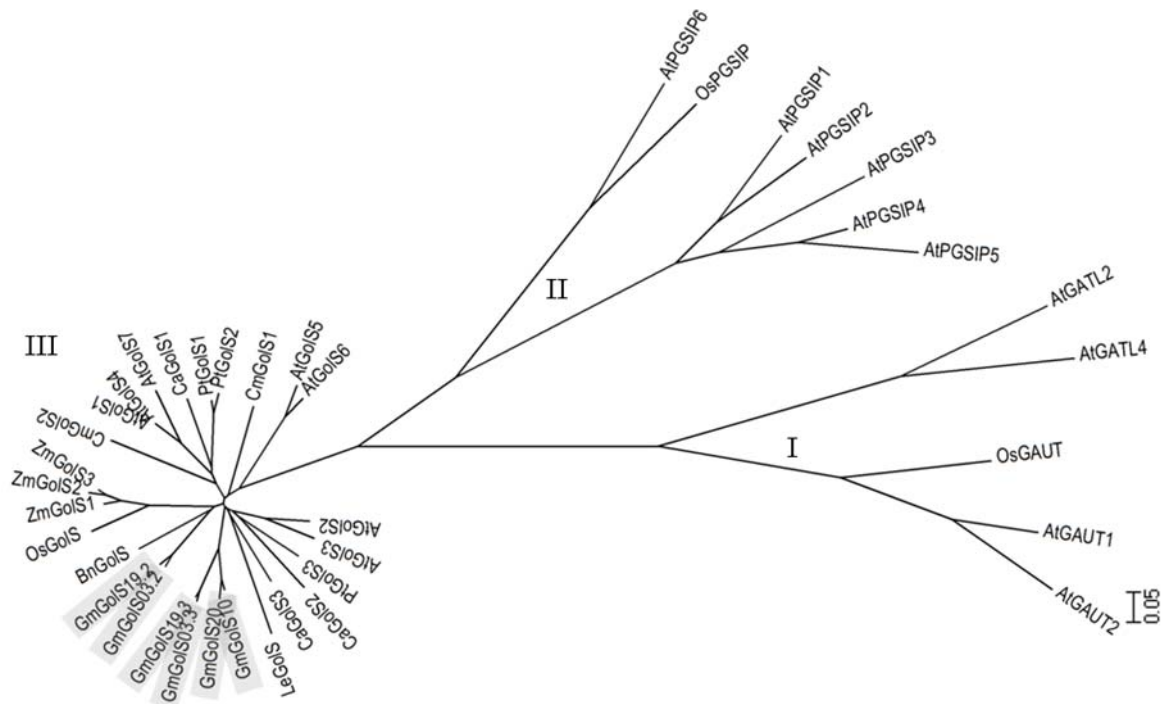


Fig. 3 Phylogenetic tree containing full-length GT8 predicted CDS from different plant families. Three main groups can be discerned. Group I, GAUT and GAUTs-like, including *Arabidopsis* (NM_115977; NM_130212; NM_114936; NM_111501) and *Oryza sativa* (LOC_Os03g11330). Group II, glycogenin-like starch initiation proteins, including *Arabidopsis* (NM_001035645; NP_177838; NM_001160812; NM_104367; NM_100770; NM_121853) and *O. sativa* (LOC_Os04g43700). Group III, plant GolSs, including *Arabidopsis* (NM_130286; NM_104537; NM_100805; NM_104734; NM_122284; NM_118758; NM_104732), *O. sativa* (LOC_Os07g48830), *Populus trichocarpa* (EU305718; EU305721; EU305723), *Coffea arabica* (GQ497218; GQ497220; GQ497220), *Brassica napus* (HM003640), *Zea mays* (AF497507; AF497508; EU967394), *Cucumis melo* (AY077642; AY077641), *Solanum lycopersicum* (AF447452), as well as the sequences homologues to six *GmGolS* putative CDS (*GmGolS03.2*, *GmGolS03.3*, *GmGolS10*, *GmGolS19.2*, *GmGolS19.3* and *GmGolS20*), indicated in gray

Initially eleven genes identified in the soybean genome were classified through the SMART and InterProscan databases as putative *RFS* or seed imbibition protein (*SIP*) genes by their conserved domains. SIPs are proteins found in seeds with unknown function. Sequence alignment has revealed a high homology between SIP and alkaline α -galactosidase (AGA) enzymes, suggesting that SIP are in fact AGA (Carmi et al. 2003, Lee et al. 2004; Hara et al. 2008). Analysis by sequence comparison showed that the alkaline enzymes (AGA or SIP) have higher similarity with RFS and STS than with acidic α -galactosidases, indicating that they belong to family 36 of GH (GH36) (Lee et al. 2004; Blöchl et al. 2008).

In our analysis, the alignment of the amino acid sequences of the eleven predicted proteins with other proteins to GH36 allowed identification of the presence of two conserved motifs (KxD and RxxD) in this family (Fig. 4), one of which has been described as a catalytic nucleophile (KxD) and the other is thought to act as a catalytic acid-base residue (RxxxD) (Li et al. 2007). The characteristics of identified soybean proteins corresponding to GH36 superfamily including the signal peptide, PI, molecular weight and some additional gene features are presented in Table 1. These results allow us to conclude that the eleven putative genes identified in soybean genome belong to the GH family 36, however they do not allow for differentiation among them.

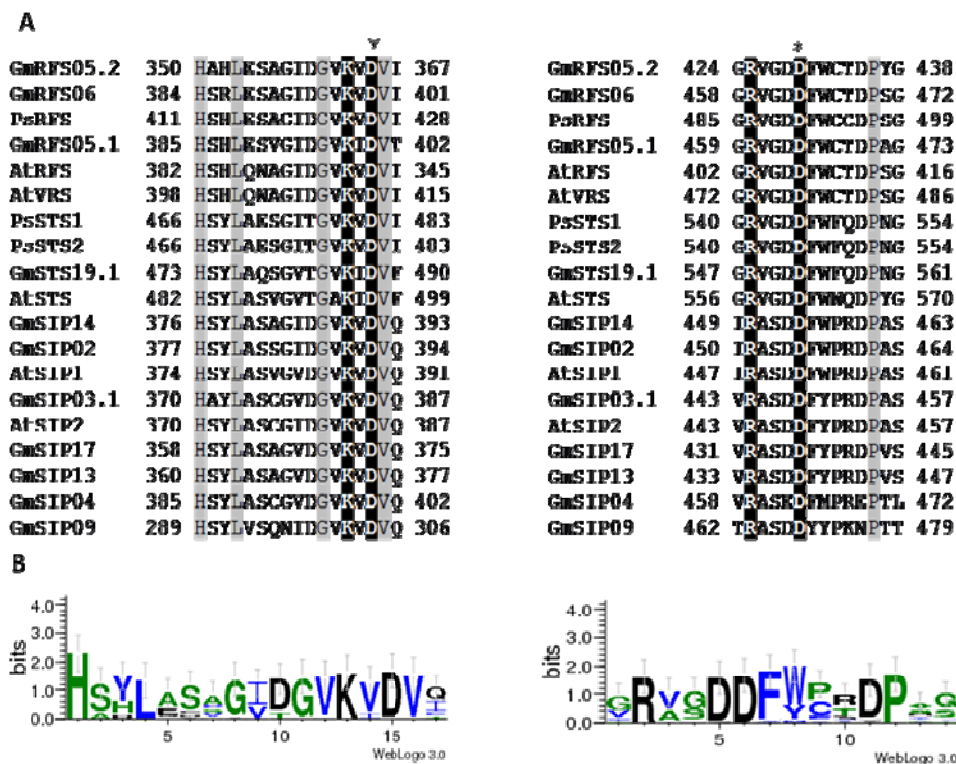


Fig. 4 Alignment of sequences of predicted RFSs, STSs and SIPs encoded by the soybean genome. (A) Boxshade output of a ClustalW sequence alignment of soybean putative RFSs, STSs and SIPs (GmSIP02; GmSIP03.1; GmSIP04; GmRFS05.1; GmRFS05.2; GmRFS06; GmSIP09; GmSIP13; GmSIP14; GmSIP17; GmSTS19.1) with other GH36 proteins from *Arabidopsis* (NP_175970; NP_191311; BAB11595; BAF00996) and *Pisum sativum* (CAD20127; CAC38094; CAD55555). Conserved motifs are highlighted in black and bold; the catalytic aspartic acid residue is marked by an arrow, and the acid-base residue with an asterisk; (B) Weblog output of the amino acid conservation of GH36 enzymes aligned in the program WebLogo 3.0 (<http://weblogo.threeplusone.com/>)

To distinguish between *SIP*, *AGA*, *RFS* and *STS* genes present in the soybean genome, we constructed a phylogenetic tree based on the predicted CDS (Fig. 5). Based on the protein sequences and the produced phylogenetic tree, SIPs, AGAs, RFSs and

STSs from plant were classified into three main groups. Group I is formed by GH36 proteins with α -galactosidase activities, including SIPs or AGAs, which cannot be distinguished phylogenetically given the high sequence homology between them (Blöchl et al. 2008). Group II and group III are composed of RFSs and STSs, respectively, with transglycosylation activities. This analysis allowed the eleven soybean genes to be classified into three phylogenetically distinct clusters. Thus, we clearly identified seven genes that encode soybean SIP or AGA (*GmSIP02*, *GmSIP03.1*, *GmSIP04*, *GmSIP09*, *GmSIP13*, *GmSIP14* and *GmSIP17*), three genes encoding RFS (*GmRFS05.1*, *GmRFS05.2* and *GmRFS05.3*) and one gene encoding STS (*GmSTS19.1*) in the soybean genome. We designated the putative identified soybean genes that code for SIP or AGA generically as *GmSIP*; similarly the putative soybean genes encoding for RFS and STS were designated as *GmRFS* and *GmSTS* genes, respectively.

The *GmSIP02*, *GmSIP03.1*, *GmSIP13*, *GmSIP14* and *GmSIP17* sequences showed high homology with other sequences of group I, and, although *GmSIP04* and *GmSIP09* were located in a separate clade, the topology of the tree and bootstrap value (96%; data not shown), indicate that this clade and group I clades are sister clades belonging to the same group (Fig. 5). *GmSIP14* and *GmSIP02* clustered with proteins described as AGA, while *GmSIP13* and *GmSIP17* are separated into a subgroup containing only SIPs. *GmSIP03.1* was grouped in a clade with other proteins described as both SIP and AGA. In fact, the relationship between SIP and AGA proteins has not been fully clarified; Carmi et al. (2003) demonstrated the functionally expressed LeSIP1 protein of tomato (*Solanum lycopersicum*) has AGA activity. Furthermore, barley (*Hordeum vulgare*) seeds developed AGA activity during imbibition, whereas the preexisting activity of acidic α -galactosidases did not change (Carmi et al. 2003; Blöchl et al. 2008).

Examination of the distribution branches in group II indicate that *GmRFS05.1* and *GmRFS06* are visibly clustered in a clade together with other RFS from western poplar (*P. trichocarpa*) (*PtRFS*), cassava (*Manihot esculenta*) (*MeRFS*), wild grape (*V. vinifera*) (*VvRFS*), *Medicago truncatula* (*MtRFS*), fava bean (*V. faba*) (*VfRFS*) and pea (*P. sativum*) (*PsRFS*), while *GmRFS05.2* formed an independent clade with *Arabidopsis* (*AtRFS*) and cucumber (*C. sativus*) (*CsRFS*).

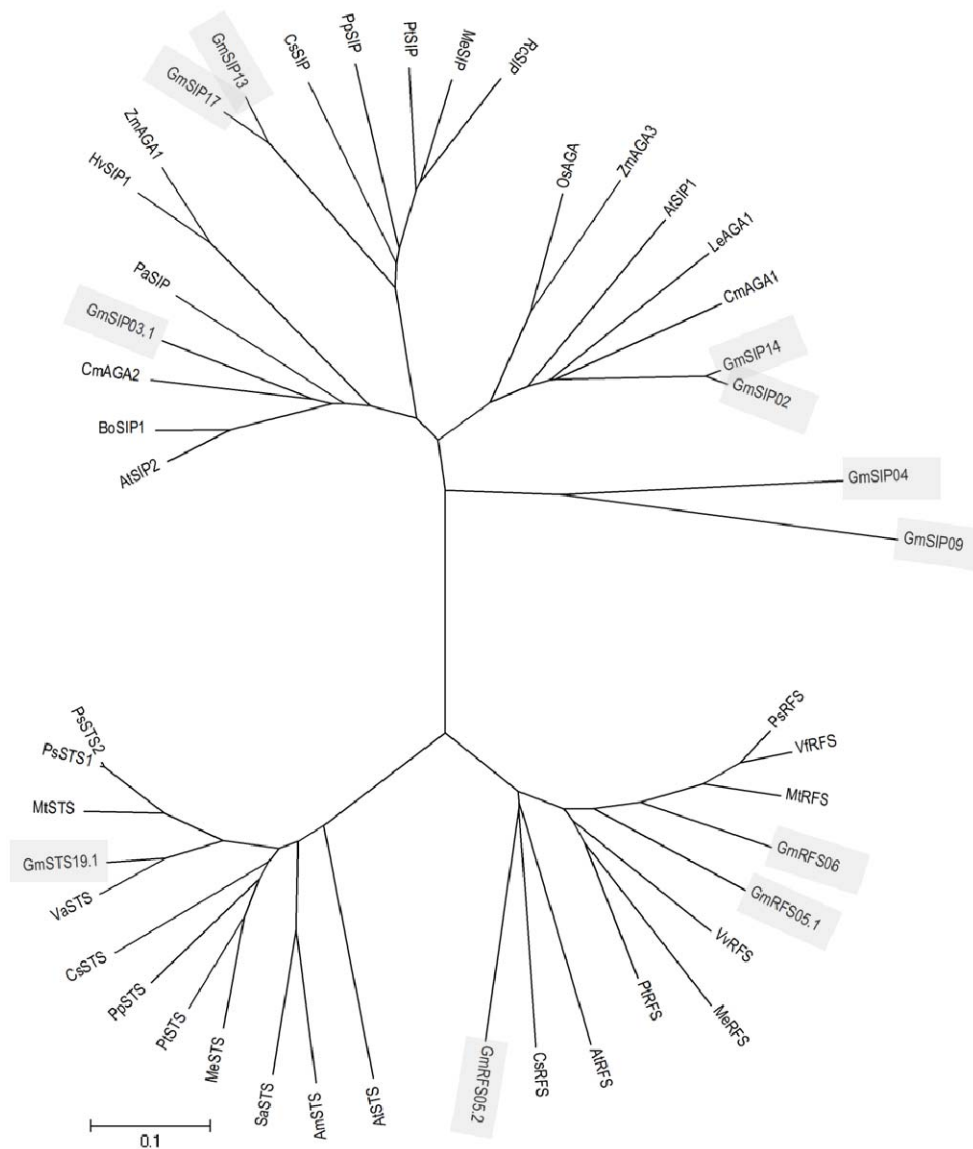


Fig. 5 Phylogenetic tree containing full-length GH36 family predicted CDS from different plant species. Group I: SIPs and AGAs, covering *Arabidopsis* (NM_104450; NM_115612), *Solanum lycopersicum* (AF512549), *Cucumis melo* (AY114164; AY114165), *Persea americana* (AJ133148), *Oryza sativa* (AF251068), *Brassica oleracea* (X79330), *Hordeum vulgare* (M77475), *Zea mays* (NM_001112323; AF497512), *Manihot esculenta* (scaffold07991), *Ricinus communis* (XM_002531912), *Populus trichocarpa* (XM_002308025), *Cucumis sativus* (scaffold00464) and *Prunus persica* (ppa001896m), as well as the CDS sequences homologues to seven *GmSIPs* putative soybean genes. Group II: RFSs, involving *Arabidopsis* (AB006702), *Cucumis sativus* (AF073744), *Pisum sativum* (AJ426475), *Vicia faba* (E24423), *M. esculenta* (cassava4.1_027087m.g), *P. trichocarpa* (XM_002321589), *Vitis vinifera* (XM_002281447), *Medicago truncatula* (Medtr3g114540), as well as the CDS sequence homologues to three *GmRFS* putative soybean gene. Group III: STS, including *P. persica* (ppa025926m.g), *P. sativum* (AJ311087; AJ512932), *Vigna angularis* (Y19024), *Alonsoa meridionalis* (AJ487030), *Stachys affinis* (AJ344091), *Arabidopsis* (AK229121), *C. sativum* (EF382356), *M. esculenta* (cassava4.1_028359m.g), *P. trichocarpa* (POPTR_0014s11370), *M. truncatula* (Medtr4g136550), as well as the CDS sequence to one *GmSTS* putative soybean gene. The putative identified soybean genes are indicated in gray

It is suggested that the soybean genome has four genes coding for RFS (Dierking and Bilyeu 2008). Our analysis indicates that there are effectively three *RFS* genes in the soybean genome: *GmRFS05.1*, *GmRFS05.2* and *GmRFS06*. These have high sequence identity with *RFSs* genes previously described (about 99%) designated *GmRFS2*, *GmRFS3* and *GmRFS4* (Dierking and Bilyeu 2008). Mutations in *RFS* genes previously described (*GmRFS2*) have been shown to result in an altered seed oligosaccharide phenotype. The resulting phenotypes were an increase in seed sucrose content as well as a decrease in both raffinose and stachyose seed oligosaccharide content (Hitz et al. 2002; Dierking and Bilyeu 2008). The fourth candidate gene, available in the Genbank as *GmRFS1* (E25448), was aligned with the genes identified to family 36 and shown to be identical to the gene sequence *GmSIP03.1*, phylogenetically classified as SIP or AGA. This evidence indicates that the gene described as *GmRFS1* does not code for a RFS, but for a SIP or AGA.

Only GmSTS19.1 was clearly classified into group III of plants STS (Fig. 5), suggesting the presence of a single gene encoding STS (*GmSTS*) in the soybean genome. Its sequence was inserted in an independent clade together with adzuki bean STS (*V. angularis*) (*VaSTS*). The alignment between these predicted soybean proteins (Supplementary Fig. 1), revealed a significant difference in putative STS. GmSTS19.1 protein contains a central region, with about 80 amino acid residues (positions 305 to 383 of GmSTS19.1), which is also present in other described STS sequences. This region has low identity considering the STS known, and therefore appears to be diagnostic for STS (Peterbauer et al. 2002). The identities between GmSTS19.1 and STS from adzuki bean (CAB64363), pea (CAC38094), *S. affinis* (CAC86963), *A. meridionalis* (CAD31704), *Arabidopsis* (NP_192106) and were 86%, 75%, 66%, 65% and 57%, respectively.

The indication of the presence of only one gene for STS in the soybean genome is not a surprise, since it has been suggested that the number of gene family members is small, not likely to be greater than two. According to Peterbauer et al. (1999), gel blot analysis with genomic DNA from adzuki bean suggested the presence of two highly homologous genes in the genome. Similarly, southern blot analysis to analyze the gene family encoding STS, suggests that the number of genes in this family is no more than two in the *A. meridionalis* genome (Voitsekhovskaja et al. 2009). In addition, the *Arabidopsis* genome seems to contain one STS gene (AK229121), based on sequence similarity with pea and adzuki bean.

Based on three different types of software, we checked the predicted subcellular localization for GmSIP, GmRFS and GmSTS predicted proteins. Except for GmSIP14 and GmSIP03, all identified proteins GmSIP were found to be cytoplasmatic, with no signal peptide. The pH optima of this enzyme class suggests cytosolic location, since it possesses a generally neutral-alkaline pH. AGA or SIPs from melon (*C. melo*) and pea are also described as cytosolic enzymes (Carmi et al. 2003; Blöchl et al. 2008). Interestingly, a rice AGA, for which no signal peptide predicted, is localized in the chloroplasts of mesophyll cells of the senescing leaves suggesting an important role for this enzyme in the degradation of chloroplast galactolipids during leaf senescence (Lee et al. 2004; Lee et al. 2009). In our analysis, the predicted subcellular location of GmSIP14 and GmSIP03.1 indicate that both proteins can also be localized in the chloroplast, even with no signal peptide, similar to the rice protein.

Exact information on the subcellular location of the raffinose and stachyose biosynthetic pathway is minimal. Our prediction of the subcellular localization suggests that both GmRFS and GmSTS are located in the cytoplasm and don't have signal peptide. With the exception of *myo*-inositol, which is synthesized directly in the chloroplast (Lackey et al. 2003), the other mentioned carbohydrates (galactinol, raffinose and stachyose) are proposed to be synthesized in the cytosol and then transported to other organelles such as chloroplasts and vacuoles where they can accumulate (Bachmann and Keller 1995; Peterbauer et al. 2001; Schneider and Keller 2009).

The mapping of the genes identified was based on the map originated by consensus sequence of the soybean genome provided by the USDA Soybase database. To our knowledge, this is the first time that genes involved in biosynthesis and intracellular content of ROs have been positioned on the physical map of soybean (Fig. 6), which can be used as an important tool in genomic studies, although molecular evidence is still required. The putative genes encoding GmGolSs were mapped on to four different chromosomes: 03, 10, 19 and 20. Chromosomes 05 and 06 anchor the genes involved in the synthesis of raffinose. The *GmSTS* gene is positioned between the microsatellite markers Satt664 and Satt229, on chromosome 19, while the putative genes coding for GmSIPs are distributed among chromosomes 02, 03, 04, 09, 13, 14 and 17. The chromosome, the linkage group, the position and the microsatellite markers flanking of each putative gene are available in Fig. 6.

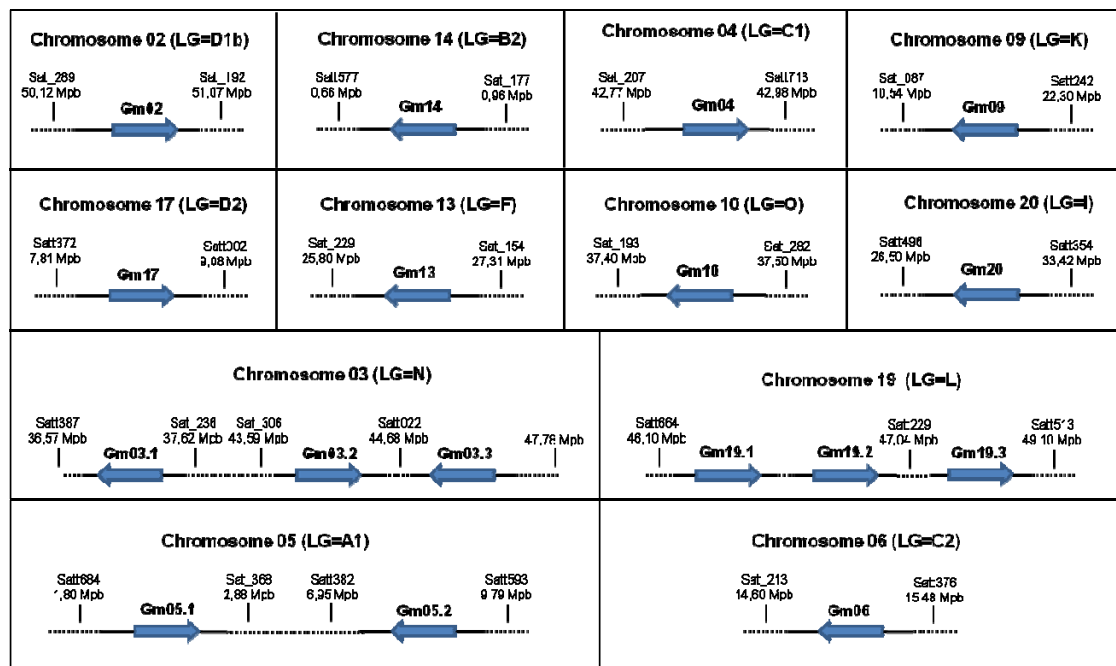


Fig. 6 Positioning of predicted genes encoding GmSIPs, GmGolSs, GmSTS and GmRFSs in soybean physical map

Understanding genome organization makes it easier to determine the exon-intron arrangement of the genome of plant species and thus to compare the genomic structure of orthologous genes. In our analysis, the results of exon-intron arrangement determination (Fig. 7) of putative genes showed that most of the *GmGolS* genes have 4 exons separated by 3 introns, except *GmGolS03.2* and *GmGolS19.2* which showed only 3 exons; both exon-intron arrangements cited were found in plant *GolS* (Taji et al. 2002; Downie et al. 2003). Five *GmSIP* genes contain 13 exons and 12 introns, in accordance with the structure of other plant SIP or AGA proteins (Carmi et al. 2003). The *GmSIP13* and *GmSIP17* genes have no introns since the available genomic sequence in phytozome database is equivalent with the CDS. *GmRFS05.1* and *GmRFS05.2* were predicted to contain 5 exons and 4 introns, confirming the structures previously described for these genes (Dierking and Bilyeu 2008). However the *GmRFS06* gene has shown to have one little more exon in the 5' end. The organization of the three putative *RFS* genes is clearly distinct from the organization of *GmSIPs*, as well as the arrangement of the putative *STS* gene. The organization of soybean *STS* gene is similar to that found in the gene structure of *Arabidopsis*. Both arrangements are characterized by 4 exons and 3 introns. Other putative plant *STS* have fewer exons, although some of the exons/intron boundary positions present in *Arabidopsis* and soybean *STS* remains conserved (Dierking and Bilyeu 2008). The comparison of the structures of putative

genes of soybean corroborate for effective existence of 6 putative *GmGolSs*, 3 predicted *GmRFSs*, 1 putative *GmSTS* and 7 candidate *GmSIPs* genes in soybean genome.

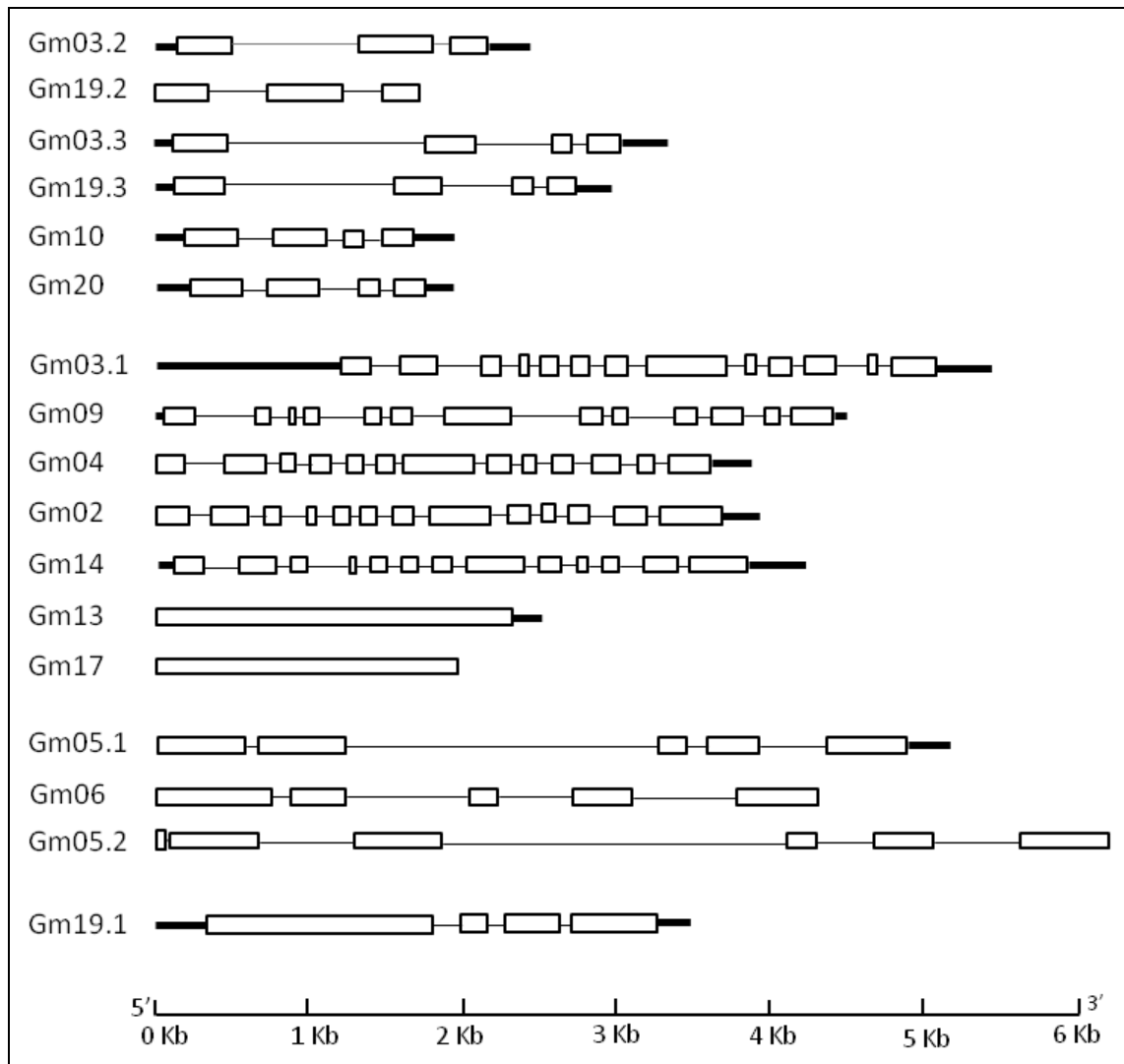


Fig. 7 The exon-intron structures of putative and known plant *GolS*, *SIP* or *AGA*, *RFS* and *STS* genes. (A) The exon-intron arrangements of soybean identified genes. Exons are represented as white boxes separated by fine lines representing introns; untranslated regions (UTR) are represented by thick grey lines.

3.2. Expression analysis

The identification of known cis-elements in the promoter regions of seventeen identified genes was performed using a sequence of 2,000 bases pairs (bp) upstream of the translation start site (+1) of the corresponding genes. The promoter regions of genes identified in soybean have at least one cis-regulatory DNA sequence involved in seed-specific expression, response to stress (biotic or abiotic) and sugar metabolism. The

names and functions of common cis-elements found into putative promoter regions of all *GmGolSs*, *GmRFSs*, *GmSTS* and *GmSIPs* are given in Table 2. Except Pyrimidine-box, motifs associated with sugar responsive or sugar repression genes, such as A-box, DOFCORE and GARE (Gibberellin Responsive Element), are found only in the 5' regulatory region of the identified genes involved in biosynthesis of oligosaccharides and raffinose, including *GmGolSs*, *GmRFSs* and *GmSTS*. This would imply that the expression of these genes may be well regulated by sugar levels. However, the sensitivity of these genes to sugar levels and more generally, the physiological significance of the different boxes identified in the 5' regulatory region will need to be confirmed by reporter gene experiments.

Three types of transcription factors regulate *Gols* genes expression in plant: HSFs (heat shock factors), DREB1A/CBF3 (dehydration-responsive element binding factor 1A/cold responsive element binding factor 3) and WRYK (involved in salicylic acid signaling and disease responses) (Taji et al. 2002; Ogawa et al. 2007; Wang et al. 2009). Usually, the binding sites for these genes are HSE (CNNGAANNTTCNNG), DRE (A/GCCGAC) and W-box (TTGAC; TGACT; TGACY; TGAC), respectively. All 5' regulatory regions of identified genes *GmGolS* contain the HSE and W-box elements, however, DRE was only present in *GmGolS10* and *GmGolS20* genes with variable frequency.

Cis-regulatory DNA sequences for seed-specific expression, such as SEF (Soybean Embryonary Factors), E-box and CAAT box, are present in all promoters of the identified genes coding for *GmSIPs*. RY Repeat and DPBFCORE (binding site for transcription factors DPBF) cis-elements are found in most promoters of *GmSIP* identified genes, except for *GmSIP03.1* and *GmSIP04*.

In the 5' regulatory regions of putative *GmRFS* and *GmSTS* genes promoter elements involved in sugar responsive or sugar repression, seed-specific expression and tolerance to biotic and abiotic stress were found. The presence of these elements agrees with studies that demonstrate the accumulation of raffinose and stachyose in response to environmental stress (Nishizawa et al. 2008; Santos et al. 2011). As stachyose is the most abundant raffinose oligosaccharide in soybean seed, it is important to emphasize the presence of these seed-specific cis-elements in the 5' regulatory region upstream of the coding regions of the *GmSTS* gene; the promoter of this gene contains RY Repeat, similar to the promoter region of *GmRFS06* gene encoding RFS.

Table 2 Common potential cis-elements identified in the 5' regulatory sequences of genes encoding for GmGolSs, GmRFSSs, GmSTSs and GmSIPs

Cis-acting elements	Genes	Sequence	Function
A-box	<i>GmGolSs; GmRFSSs; GmSTS</i>	TACGTA	Sugar repression
ABRE	<i>GmRFSSs; GmSTS; GmSIPs</i>	(C/1)ACG(T/C)G(T/C/G)	Abcisic acid responsive element; a cis-acting regulatory element involved in abcisic acid responsiveness
CAAT Box	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	CAAT	Required for promoter activity seed specificity
Cure	<i>GmGolSs; GmRFSSs; GmSTS</i>	GTAC	Copper responsive element and also involved in oxygen response
DOFCORE	<i>GmGolSs; GmRFSSs; GmSTS</i>	AAAG	Binding site for DOF transcript factors involved in carbon metabolism and induced by salicylic acid
DPBFCORE	<i>GmGolSs; GmRFSSs</i>	ACACNNG	Binding elements of transcription factors DPBF (Dc3 promoter-binding factor) 1 and 2, involved in embryo-specific transcription
E-box	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	CANNTG	Essential element for the expression of storage proteins
Erd1	<i>GmGolSs; GmRFSSs; GmSTS</i>	ACGT	Cis-acting regulatory element required for early response to dehydration (etiolation-induced expression)
GARE	<i>GmGolSs; GmRFSSs; GmSTS</i>	TAACAA(G/A)	Gibberellin Responsive Element; a cis-acting regulatory element involved in Gibberellin responsiveness and also partially involved in sugar repression
GATA	<i>GmGolSs; GmRFSSs; GmSIPs</i>	GATA	Cis-acting regulatory element required for high level light regulated and tissue specific expression
G-box	<i>GmRFSSs; GmSIPs</i>	CACGTG	Cis-acting regulatory element involved in light responsiveness
GT-1	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	GAAAAA; GGTTAA	Cis-acting regulatory element required for rapid response to pathogen attack, salinity and salicylic acid inducible gene expression
HSE	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	CNNGAANNTTCNNG	Heat Stress Element: a cis-acting regulatory element involved in Heat stress responsiveness
I-box	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	GATAA	Light box Element: a cis-acting regulatory element conserved in sequence upstream of light-regulated genes
MART-box	<i>GmGolSs; GmRFSSs; GmSTS</i>	TTWTWTTWTT	Matrix attachment region - acts as an enhancer in constructs for expression of transgenes
MYB	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	WAACCA; AACTG; CNGTTR; YAACKG; GGATA; CAACTG	Cis-acting regulatory element involved in regulation of drought inducible gene expression
MYC	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	CATGTG; CACATG; CANNTG	Cis-acting regulatory element involved in regulation of drought inducible gene expression
Pyrimidine box	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	TTTTTCC; CCTTTT	Partially involved in sugar repression. It requires Gibberellin for its induction
SEF	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	ATATTTAWW; AACCCA; RTTTTTR;	Enhancer present in the promoters in the genes of β -conglycinin in soybean
TATA-box	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	TATAAAT	Positionally important element for the accuracy of the transcription start
W-box	<i>GmGolSs; GmRFSSs; GmSTS; GmSIPs</i>	TTGAC; TGACT; TGACY; TGAC	Cis-acting regulatory element involved in direct fungal elicitor stimulated transcription of defense genes and activation of genes involved in response to wounding

Tissue-specific gene expression is one alternative for a subset of duplicated soybean genes (Chappell et al. 2006). To assist in understanding the roles of the soybeans genes identified and to determine if different soybean isoforms (GmGolSs,

GmRFSs, GmSTSSs and GmSIPs) are functionally similar or different, the virtual expression profile, equivalent to virtual northern blot analysis (VNB), of seventeen identified genes was determined. We specifically examined the EST frequency of different soybean genes identified and only those EST which had E-value $<1e^{-100}$ were considered for counting the number of hits. Redundant ESTs were removed manually.

The expression pattern of *GmGolSs* in different organs is shown in Fig. 8. *GmGolS* transcripts appear to be accumulated in soybean seeds. In accordance with the osmoprotectant role attributed to the ROs, the presence of *GolS* transcripts in plant seeds was expected and has been previously described in tomato (Downie et al. 2003), melon (Volk et al. 2003) and corn (Zhao et al. 2004). Zhao et al. (2004) described a differential pattern of gene expression for two isoforms of *ZmGolS*. *ZmGolS2* and *ZmGolS3* gene family members were detected in seeds, however, *ZmGolS2* was detected only in response to stress. The third member of this family was not detected in the tissue examined. This experimentally determined pattern of differential expression displayed by the *ZmGolS* genes was reinforced by expression analysis *in silico*. Expression analysis also indicates that transcripts of *GmGolS* genes seem also to accumulate in leaves and roots of plants submitted to biotic and abiotic stresses (Fig. 8). The expression of genes *GolS* has been widely described and associated with tolerance to stress in different plant species (Taji et al. 2002; Chen et al. 2002; Zhao et al. 2004, Nishizawa et al. 2008; Santos et al. 2011); however, not all of the *GolS* genes within a species are up-regulated during plant stress. Of the seven genes identified in *Arabidopsis*, only three were shown to be expressed in stressed plants, while in corn, the *ZmGolS3* gene appears to be expressed only during seed development (Taji et al. 2002, Zhao et al. 2004).

The *GmSIPs* transcripts displayed a significant prevalence in seeds as compared to other organs analyzed (Fig. 8). The *AGA* gene family is highly expressed in germinating seeds for breakdown of oligosaccharides (Carmi et al. 2003; Zhao et al. 2006; Blöchl et al. 2008); in seeds of corn and barley the alkaline enzymes were either already present in the dry seed (Zhao et al. 2006) or expressed at early stages of germination (Carmi et al. 2003). In addition, tissue prints of germinating pea seeds and real time PCR revealed that the *PsAGA1* transcripts were present in equal abundance in the embryo and cotyledons (Blöchl et al. 2008). *AGA* transcript accumulation can also be regulated by abiotic stress, although the significance of this process is not well understood. The expression of *AGA* genes was up-regulated by different abiotic stress in

corn (Zhao et al. 2006), rice (Lee et al. 2004) and *Tetragonia tetragonioides* (Hara et al. 2008). In contrast, a maize *ZmAGAI* transcript accumulated when seed germination was interrupted by abiotic stresses. Some *GmSIPs* genes (*GmSIP17*, *GmSIP13*, *GmSIP03.1* and *GmSIP14*) were shown to be expressed in callus. This expression pattern was also reported in corn for the gene *ZmAGAI* demonstrating that transcripts accumulate in callus and embryo-derived as well as seed (Zhao et al. 2006). No EST significantly similar to *GmSIP09* gene was found in the database library ESTs, whereas the E-value $<1e^{-100}$.

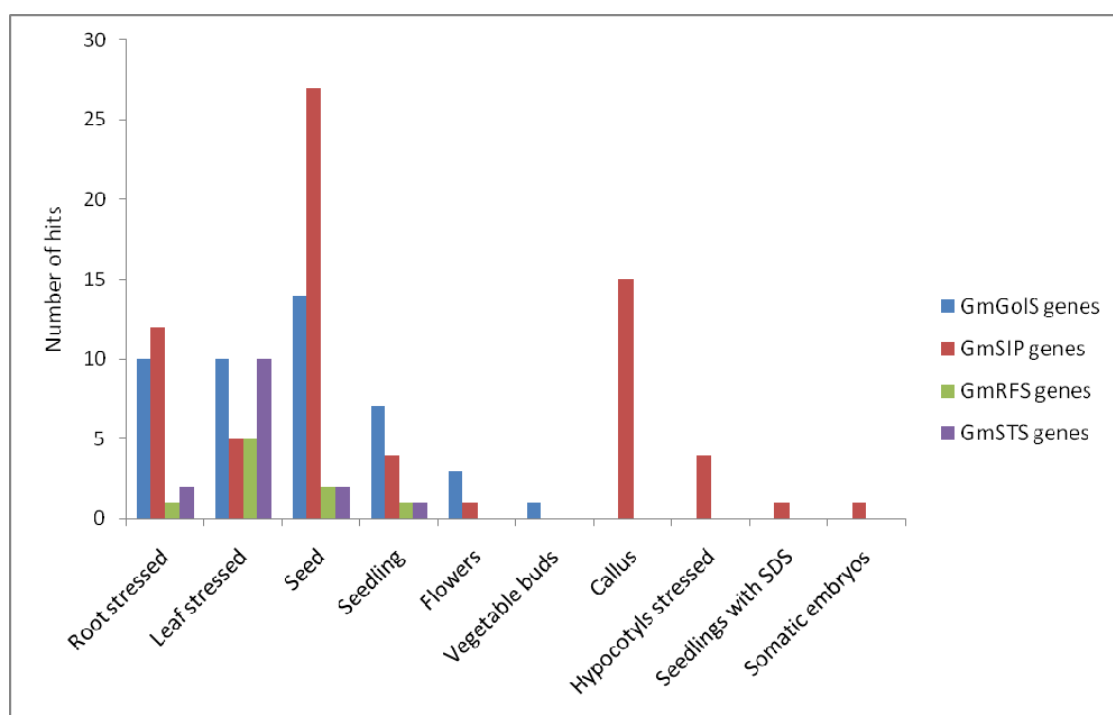


Fig. 8 *In silico* analysis of tissue-specific expression of genes *GmGolS*, *GmSIPs*, *GmRFSs* and *GmSTS* from a database of 1,461,623 soybean ESTs

Our *in silico* gene expression analysis indicated that *GmRFSs* genes can be up-regulated in response to biotic and abiotic, but can also be expressed in seeds and seedling (Fig 8). In a previous study, Dierking and Bilyeu (2008) verified, by real time PCR, that transcript level of two putative *GmRFSs* genes are similar in different organs (leaves, seedlings and seeds), and expression of this two genes appears no reduce in tissues of seed development. *RFS* transcripts have been shown to accumulate in response to abiotic stress in *Arabidopsis*, mainly in leaves (Nishizawa et al. 2008). However, similar to *GolS* genes, not all *RFS* genes described in *Arabidopsis* are involved in stress tolerance; under oxidative damage caused by treatment with

methylviologen (MV), to simulate oxidative stress, the expression of four of six putative *AtRFS* genes was induced in leaves, resulting in increased raffinose levels.

Although *GmSTS19.1* transcripts were present in seedlings, seeds and stressed roots, their accumulation was significantly higher in stressed leaves (Fig 8). In a previous analysis conducted by our group, the transcript levels of *GmSTS19.1* in soybean were quantified by semi-quantitative RT-PCR. This analysis revealed that *GmSTS19.1* is expressed in seed (throughout development), stem, root, and predominantly in leaves (Fialho 2007; data not shown), strengthening our *in silico* expression analysis. The presence of STS transcripts in pea seeds was detected during mid- to late stages of development, but not in germinated seeds, indicating that in this species ROs can provide an important carbon source for germination (Peterbauer et al. 1999). Although expression of *STS* genes in response to physiological stress is poorly described, the significant accumulation of *GmSTS19.1* mRNA in stressed leaves of soybean indicated by our analysis, suggests that stachyose may act as an osmoprotectant in soybean during biotic and abiotic stresses. Although this carbohydrate does not accumulate in *Arabidopsis* leaves in response to several types of stress (Taji et al. 2002; Nishizawa et al. 2008), the stachyose levels in coffee (*C. arabica*) leaves were much higher than galactinol and raffinose after water deficit imposition (Santos et al. 2011), reinforcing its possible role in protection against damage caused mainly by drought stress.

The expression of the *GmSTS19.1* gene, mainly in soybean leaves can also be associated with a subsequent translocation of sugar to the other organs of plants, such as seed, as a source of energy. In Cucurbitaceae, Scrophulariaceae and other plant families, ROs act in carbohydrate transport, where they are not accumulated in the phloem but efficiently transported from the leaves to other parts of plants (Ayre et al. 2003; Voitsekhovskaja et al. 2009). Recently, a study with *A. meridionalis* suggested that in this species, which translocates ROs in addition to sucrose, the expression of *STS* gene occur in intermediate cells of leaf minor veins and the promoter *AmSTS* is active not only in leaf but also in cotyledons (Voitsekhovskaja et al. 2009). Further study is required to determine if stachyose is translocated from one organ to another in soybean, similar to that which occurs in some plant species already characterized, and how often this carbohydrate is translocated.

In summary, in this study we have identified six putative genes encoding GolSs, three putative genes encoding RFSs and one putative gene encoding STS. To our

knowledge, this is the first report of *in silico* identification and characterization of *STS* gene in soybean. Additionally, we also identified and characterized *in silico* seven putative genes coding for SIPs or AGAs. Through sequence identity and phylogenetic analysis, by homology, it can be inferred that the predicted soybean GmGolS, GmRFS and GmSTS proteins are possibly involved in the biosynthesis of ROs, while the GmSIPs proteins may be associated with degradation of these oligosaccharides. Our results together indicate that the putative identified genes may be involved in desiccation tolerance during seed maturation processes and in response to biotic and abiotic stresses, directing further experimental studies for confirmation.

4. REFERENCES

- Ayre BG, Keller F, Turgeon R (2003) Symplastic continuity between companion cells and the translocatin stream: Long-distance transport is controlled by retention and retrieval mechanisms in the phloem. *Plant Physiol* 131:1518-1528. doi: 10.1104/pp.012054
- Bachmann M, Keller F (1995) Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L.: inter- and intracellular compartmentation. *Plant Physiol* 109:991–998. doi: 10.1104/pp.109.3.991
- Blöchl A, Peterbauer T, Hofmann J, Richter A (2008) Enzymatic breakdown of raffinose oligosaccharides in pea seeds. *Planta* 228:99–110. doi: 10.1007/s00425-008-0722-4
- Carmi N, Zhang G, Petreikov M, Gao Z, Eyal Y, Granot D, SchaVer AA (2003) Cloning and functional expression of alkaline α -galactosidase from melon fruit: similarity to plant SIP proteins uncovers a novel family of plant glycosyl hydrolases. *Plant J* 33:97–106. doi: 10.1046/j.1365-313X.2003.01609.x
- Chappell AS, Scaboo AM, Wu X, Nguyen H, Pantalone VR, Bilyeu KD (2006) Characterization of the MIPS gene family in *Glycine max*. *Plant Breed* 125:493–500. doi: 10.1111/j.1439-0523.2006.01264.x
- Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budwortha PR, Tsoa Y, Xie Z, Chene X, Lame S, Krepsa JA, Harperf JF, Si-Ammourc A, Mauch-Manic B, Heinleing M, Kobayashig K, Hohng T, Dangl JL, Wanga X, Zhu T (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14:559–574. doi: 10.1105/tpc.010410

- Dey PM (1985) D-Galactose-containing oligosaccharides. In: Dey PM and Dixon RA (eds) *Biochemistry of Storage Carbohydrates in Green Plants*. Academic Press, New York, pp 53-129.
- Dierking EC, Bilyeu KD (2008) Association of a Soybean Raffinose Synthase Gene with Low Raffinose and Stachyose Seed Phenotype. *Plant Genome* 2:135-145. doi: 10.3835/plantgenome2008.06.0321
- Downie B, Gurusinghe S, Dahal P, Thacker RR, Snyder JC, Nonogaki H, Yim K, Fukunaga K, Alvarado V, Bradford KJ (2003) Expression of a GALACTINOL SYNTHASE Gene in Tomato Seeds is Up-Regulated before Maturation Desiccation and Again after Imbibition whenever Radicle Protrusion is Prevented. *Plant Physiol* 131:1347–1359. doi/10.1104/pp.016386
- Fialho LS (2007) Purificação e caracterização de uma α -galactosidase em sementes de *Tachigali multijuga* e clonagem parcial do gene da estaquiase sintase de soja. Dissertation, Universidade Federal de Viçosa.
- Grant D, Nelson RT, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucl Acids Res* 38:D843-D846. doi: 10.1093/nar/gkp798
- Haagenson DM, Klotz KL, Campbell L (2008) Impact of storage temperature, storage duration, and harvest date on sugarbeet raffinose metabolism. *Postharvest Biol Technol* 49:221–228. doi:10.1016/j.postharvbio.2008.02.007
- Hara M, Tokunaga K, Kuboi T (2008) Isolation of a drought-responsive alkaline α -galactosidase gene from New Zealand spinach. *Plant Biotechnol* 25:497–501.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. *Nucl Acids Res* 27:297-300. doi: 10.1093/nar/27.1.297
- Hitz WD, Carlson TJ, Kerr PS, Sebastian SA (2002) Biochemical and Molecular Characterization of a Mutation That Confers a Decreased Raffinosaccharide and Phytic Acid Phenotype on Soybean Seeds. *Plant Physiol* 128:650-660. doi: 10.1104/pp.010585
- Lackey KH, Pope PM, Johnson MD (2003) Expression of 1L-myoinositol-1-phosphate synthase in organelles. *Plant Physiol* 132:2240–2247. doi: 10.1104/pp.103.020610
- Lee RH, Hsu JH, Huang HJ, Lo SF, Chen SC (2009) Alkaline alpha-galactosidase degrades thylakoid membranes in the chloroplast during leaf senescence in rice. *New Phytol*. 184:596–606. doi: 10.1111/j.1469-8137.2009.02999.x

- Lee RH, Lin MC, Chen SCG (2004) A novel alkaline α -galactosidase gene is involved in rice leaf senescence. *Plant Mol Biol* 55:281–295. doi: 10.1007/s11103-004-0641-0
- Lehle L, Tanner W (1973) The function of myo-inositol in the biosynthesis of raffinose. Purification and characterization of galactinol:sucrose 6-galactosyltransferase from *Vicia faba* seeds. *Eur J Biochem* 38:103–110. doi:10.1111/j.1432-1033.1973.tb03039.x
- Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P (2004) SMART 4.0: towards genomic data integration. *Nucl Acids Res* 32:142-144. doi: 10.1093/nar/gkh088
- Li S, Li T, Kim W-D, Kitaoka M, Yoshida S, Nakajima M, Kobayashi H (2007) Characterization of raffinose synthase from rice (*Oryza sativa* L. var. Nipponbare). *Biotechnol Lett* 29:635–640. doi: 10.1007/s10529-006-9268-3
- Liu J-J, Odegard W, De Lumen BO (1995) Galactinol synthase from kidney bean cotyledon and zucchini leaf. *Plant Physiol* 109: 505–511.
- Lowell CA, Kuo TM (1989) Oligosaccharide metabolism and accumulation in developing soybean seeds. *Crop Sci* 29:459-465. doi:10.2135/cropsci1989.0011183X002900020044x
- Nishizawa A, Yabuta Y, Shigeoka S (2008) Galactinol and Raffinose Constitute a Novel Function to Protect Plants from Oxidative Damage. *Plant Physiol* 147:1251–1263. doi: 10.1104/pp.108.122465
- Obendorf RL (1997) Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Sci Res* 7:63–74. doi: 10.1017/S096025850000341x
- Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S (2003) Gibberellin biosynthesis and response during *Arabidopsis* Seed germination. *Plant Cell* 15:1591–1604. doi: 10.1105/tpc.011650
- Peterbauer T, Mucha J, Mach L, Richter A (2002) Chain-enlongation of raffinose in pea seeds. Isolation, characterization, and molecular cloning of multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. *J Biol Chem* 277:194–200. doi: 10.1074/jbc.M109734200
- Peterbauer T, Mucha J, Mayer U, Popp M, Glossl J, Richter A (1999) Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*): molecular cloning and functional expression of stachyose synthase. *Plant J* 20:509-518. doi: 10.1046/j.1365-313X.1999.00618.x

- Peterbauer T, Richter A (1998) Galactosylononitol and stachyose synthesis in seeds of adzuki bean. Purification and characterization of stachyose synthase. *Plant Physiol* 117:165-172. doi: 10.1104/pp.117.1.165
- Peterbauer T, Richter A (2001) Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Sci Res* 11:185-197. doi: 10.1079/SSR200175
- Prestridge DS (1991) SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. *Comput Appl Biosci* 7:203-206. doi: 10.1093/bioinformatics/7.2.203
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Santos TB, Budzinski IGF, Marur CJ, Petkowicz CLO, Pereira LFP, Vieira LGE (2011) Expression of three galactinol synthase isoforms in *Coffea arabica* L. and accumulation of raffinose and stachyose in response to abiotic stresses. *Plant Physiol Biochem* doi:10.1016/j.plaphy.2011.01.023.
- Saravitz DM, Pharr DM, Carter TE (1987) Galactinol synthase activity and soluble sugars in developing seeds of four soybean genotypes. *Plant Physiol* 83:185–189. doi: 10.1104/pp.83.1.185
- Schneider T, Keller F (2009) Raffinose in Chloroplasts is Synthesized in Cytosol and Transported across the Chloroplast Envelope. *Plant Cell Physiol* 50:2174-2189. doi: 10.1093/pcp/pcp151
- Schweizer TF, Horman I, Wiirsch P (1978) Low molecular weight carbohydrates from leguminous seeds; a new disaccharide: galactopinitol. *J Sci Food Agric* 29:148-154. doi: 10.1002/jsfa.2740290212
- Senser M, Kandler O (1967) Vorkommen und verbreitung von galactinol in blättern höherer pflanzen. *Phytochemistry* 6:1533-15410. doi:10.1016/S0031-9422(00)82947-7
- Sprenger N, Keller F (2000) Allocation of raffinose family oligosaccharides to transport and storage pools in *Ajuga reptans*: the roles of two distinct galactinol synthases. *Plant J* 21:249–258. doi: 10.1046/j.1365-313x.2000.00671.x
- Taji T, Ohsumi C, Luchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki M, Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. *Plant J* 4:417-426. doi: 10.1046/j.0960-7412.2001.01227.x

- Takanashi K, Shitan N, Sugiyama A, Kamimoto Y, Hamamoto M, Iwaki T, Takegawa T, Yazaki K (2008). Galactinol Synthase Gene of *Coptis japonica* Is Involved in Berberine Tolerance. *Biosci Biotechnol Biochem* 72:398-405. doi:10.1271/bbb.70495
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599. doi: 10.1093/molbev/msm092
- Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl Acids Res* 22:4673–4680. doi: 10.1093/nar/22.22.4673
- Voitsekhovskaja OV, Rudashevskaya EL, Demchenko KN, Pakhomova MV, Batashev DR, Gamalei YV, Lohaus G, Pawlowski K (2009). Evidence for functional heterogeneity of sieve element–companion cell complexes in minor vein phloem of *Alonsoa meridionalis*. *J Exp Bot* 60:1873-1883. doi: 10.1093/jxb/erp074
- Volk GM, Haritatos EE, Turgeon R (2003) Galactinol synthase gene expression in melon. *J Amer Soc Hort Sci* 128:8–15.
- Wang Z, Zhu Y, Wang L, Liu X, Liu Y, Phillips J, Deng X (2009) A WRKY transcription factor participates in dehydration tolerance in *Boea hygrometrica* by binding to the W-box elements of the galactinol synthase (*BhGolS1*) promoter. *Planta* 230:1155–1166. doi: 10.1007/s00425-009-1014-3
- Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y (2010) Evolution and Function of the Plant Cell Wall Synthesis- Related Glycosyltransferase Family 8. *Plant Physiol* 153:1729-1746. doi:10.1104/pp.110.154229
- Zhao TY, Corum JW, Mullen J, Meeley RB, Helentjaris T, Martin D, Downie B (2006) An alkaline alpha-galactosidase transcript is present in maize seeds and cultured embryo cells, and accumulates during stress. *Seed Sci Res* 16:107–121. doi: 10.1079/SSR2006243
- Zhao T-Y, Meeley RB, Downie B (2003) Aberrant processing of a Maize GALACTINOL SYNTHASE transcript is caused by heat stress. *Plant Sci* 165:245–256. doi:10.1016/S0168-9452(03)00168-7
- Zhao T-Y, Thacker R, Corum JW, Snyder JC, Meeley RB, Obendorf RL, Downie B. (2004) Expression of the maize Galactinol Synthase gene family: (I) Expression of

two different genes during seed development and germination. *Physiol Plant*
121:634-646. doi: 10.1111/j.1399-3054.2004.00367.x

SUPPLEMENTARY

PsRFS -----MAPPSITKTATQQD-----VISTVDIGNSPLLSISLDQSRNFLVNGHPFLTQVPPNITTTTST 59
 GmRFS06 -----MAP-SISKTVELN-----SFGLVNGSNFLANGHPFLTEVPENIIVTPSP- 43
 PtRFS MGLVDGEQP-----LSITLE-GKNFLANGHPVLTEVPTNI IATPSP- 40
 GmRFS05.1 -----MAPSLTAKQEALL-----N-----ANPHLSIKLQ-SSTFFANNHPILTQVPPNITTTTTP- 50
 GmRFS05.2 -----FSLC-NSTLVKNGQVILSQVPKNVTLTPCT- 29
 PsSTS1 -----MAPPLN-----STSNLIKTESIFDLSEKFKVKG-FPLFHDVPEVNSVFRSFS 48
 MtSTS -----MAPPN-----TTLDIVKTESLDDLSEKFTVVRG-VPLFHDVPEVNSVFRSFS 46
 VaSTS -----MAPPNP-----VNATLGLPESEKVFVLDSDGKLTVKG-VVLLSHVPEVNTVFRSFS 50
 GmSTS19.1 -----MAPPNP-----VNSTLGGKSLKVFVLDSDGKFTVVRG-VPLLSQVPPNVTFRSFS 50
 SaSTS -----MAPPN-----PISSIFSP LISVKKDNAPFELVGGKLSVKN-VPLLSEIPSNVTFKSFSS 53
 AmSTS -----MAPPYDPIPIP----IPMSAILNFLSSTVKDNSFELLDGTLVKN-VPILTDIPSNVFRSFS 59
 AtSTS -----MAPLHESLSSINDVIESKPLFVPI TKPILQPNSFNLSEGLCAKDSTPILFDVQNVTFTFSS 64

PsRFS PSPFLDFKSNKD----TIANNNTLQQQCGVGFNTTEAKSHHVPLGKLGKIKFTSIFRFRKVVWTTTHW 125
 GmRFS06 ----IDAKSSK-----NNEDDDVVGCFVGFHADEPRSRHVASLGLKLRGKFMISIFRFRKVVWTTTHW 100
 PtRFS -----FLSSN-----KTKNLVGCFCVGFDAHEPKSHHVPIGKLSGIRFMSIFRFRKVVWTTTHWI 93
 GmRFS05.1 -----PHD-----ASTTPAGCFVGFSADEARSRHVISLGLKLRGIRFTSIFRFRKLWSTHWS 101
 GmRFS05.2 -----YDHTTGCFGLGFHATSPKSRHVAPLQGLKNI SFTSIFRFRKVVWTTTLWT 77
 PsSTS1 -----ICKPSESNAAPPSSLQKVLAYSHKGGFFGFSHETPSDRLMNSIGSFGKDFLSIFRFRKTWWSTQWI 113
 MtSTS -----ICKPSESNAAPPSSLVDRVLSFHKGGFFGFSGNETPSDRFMNSLGSFGRNFVLSIFRFRKTWWSTQWI 111
 VaSTS -----ICVPRD--APSSILQRVTAASHKGGFLGFSHVSPSRLINSLGSFGRNFVLSIFRFRKTWWSTQWV 113
 GmSTS19.1 -----ICEPRD--APSSILQRVIAVSHKGGFFGFSSQVSPSRLTNSLGSFGRNFVLSIFRFRKTWWSTQWV 113
 SaSTS -----ICQSSG--APAPLYNRAQSLSNCGGFLGFSSQKESADSVTNSLKGFTNREFVLSIFRFRKTWWSTQWV 116
 AmSTS -----IVQSSE--APVPLFQRAQSLSSSGFLGFSSQNEPSSRLMNSLKGFTDRDFVLSIFRFRKTWWSTQWV 122
 AtSTS -----HSISTD--APLPILLRVQANAHKGGFLGFTKESPSDRLTNSLGRFEDRELSLFRFRKMWWSTAWI 127

PsRFS LHLSNDPYRLVKEAVKVIQTKLGTFTKLEEGTNGHELQHETQILILDKNISLGRPYVLLPLILENSFRTS 195
 GmRFS06 VHVGHDPYQLREATKVVRMHLGTFKLLEEGSNHGELEHETQMMLLDKNQDQGRPFVLLPLIQASFRAS 170
 PtRFS MHVGDDPYSLVKEAMKVI RVLHGTFRLLLEEGNSKDVHEHETQIMILDRND-LGRPYVLLPLILEGPF 162
 GmRFS05.1 MHVHDDPFTLIDEALKVIRVYLGTFRLMEEGSNRDRVENETQMMILQNDAVEGRPYVLLPLILEGPF 171
 GmRFS05.2 LHAGDNPFTLVKEAMRVVRAHLGSKLLEEGSNRDLLETQFLMLQSHP-----YVLFPLILQPPFRAS 142
 PsSTS1 VHFSENPYDLMKEAYSALRVHLNFRLLLEEGSGSDLQMETQWILIEVPETKS--YVVIIP IIEKCFRSA 181
 MtSTS VHFSENPYDLMKEAYSALRVHLNFRLLLEEGSGSDLQMETQWILIEVPEIRS--YAVIIP IIEGFRSA 179
 VaSTS VHVAEPTYNLMREAYSALRVHLNFRLLLEEGSGSDLQMETQWILIEVPETES--YVVIIP IIEKCFRSA 181
 GmSTS19.1 VHVSENPNVMKEAYSALRVHLNFRLLLEEGSGSDLQMETQWVILIEVPEIKS--YVVIIP IIEKCFRSA 181
 SaSTS VHVSDNPYTLMDGYTAVRVHLDTFKLIEEGTTSQSDIQMETQWIMLNLPEIKS--YAVVIVPEVGFSA 184
 AmSTS VHVSENPTYLMRDAYTAVRVHLNFRLLLEEGTTSQSDIQMETQWIMLNDVPEIKS--YAVVIVPEVGFSA 190
 AtSTS IHICDNPYNLMKEAFSALRVHMNTFKLLEEGSGSDLQMETQWVMLKIPEIDS--YVVIIP IIEGAFRAS 195

PsRFS LQPGLNDYVDMSESGSTHVTGSTFKACLYKTPPSIIEKFGWCTWDAFYLLKVHPKGVWEGVKALTDGGCP 265
 GmRFS06 LQPGLDDYVDMSESGSTRVCGSSFGSCLYKTA PVIIDKFGWCTWDAFYLLKVHPKGVWEGVKLVGEGCP 240
 PtRFS LQPGVNDNVDICVSESGSQVCGSSFRSCLYKTPPGIVDKFGWCTWDAFYLLTVHPKGVREGVKLVGEGCP 232
 GmRFS05.1 LQPGLHDDVDICMESGARVTKSRFTSVYKTVPGIIDKFGWCTWDAFYLLNVHPKGVREGVKLVGEGCP 241
 GmRFS05.2 LQPHSDDNVAVCVSESGSHVTASSFDTVVYKTVPGIMVDKFGWCTWDAFYLLTVHPKGVREGVKLVGEGCP 212
 PsSTS1 LFPGFNDHVKIIAESGSTKVKESTFNSIAYKTLIPNLVDKFGWCTWDAFYLLTVNPIGVFHLDDFSKGGVE 251
 MtSTS LHPGSDDHVMICAESGSTKVKALSFAIAYKTLIPNLVDKFGWCTWDAFYLLTVNPIGVFHLDDFSKGGVE 249
 VaSTS LHPGSDDHVMICAESGSTQVRASSFGAIAYKTVPRIVDKFGWCTWDAFYLLTVNPIGVVHGLKDFSEGGVA 251
 GmSTS19.1 LHPGSDDHVMICAESGSTQVKASSFGAIAYKTVPKIADKFGWCTWDAFYLLTVNPIGVVHGLKDFSEGGVA 251
 SaSTS LFPKGDGHVILISAESGSTCVKTTSTFSAIYKSAEPLVKNKFGWCTWDAFYLLTVPEPAGIYHGVQEFADGGLT 254
 AmSTS LFPKGDGHVILISAESGSTKVKTSNFDIAIYKSAPPLVKNKFGWCTWDAFYLLTVPEPAGIYHGVQEFADGGLT 260
 AtSTS LTPGKGNVILICAESGSTKVKESSEFKSIAYKTLKPKIIVDKFGWCTWDAFYLLTVPEPAGIYHGVQEFADGGLT 265

PsRFS EGFVIIDDGWQSI SHDDDPVTERDGMNRTSAGEQMPCLRIKYEENYKFRYENG----- 320
 GmRFS06 EGMVLIDDGWQAICHDED-PITDQEGMKRTSAGEQMPCLRVKLEENYKFRYQYCSG----- 294
 PtRFS EGMVLIDDGWQSI CHDDD-PISEQEGMNRTAAGEQMPCLRVKFEENYKFRDYESP----- 286
 GmRFS05.1 EPLVLIDDGWQTFRCRDE-TVSDGGLNCSVPGEQMLNRLIKFEENKFKFEYKCG----- 295
 GmRFS05.2 EGFVLIDDGWQCSHSDSD--PEKEGMNQTVAGEQMPCLRIKYEENYKFRYSYK----- 262
 PsSTS1 PRFVIIDDGWQSI SFDGY--DPNEDAKNLVVGGEQMSGRLHRFDECYKFRKYESGLLLGPNSPFPDPNPF 304
 MtSTS PRFVVIDDGWQSI ILLGD--DPNVDAKNLVVGGEQMTARLHRLDEGDKFKKYESGLVLGPNSPFPNPKTI 317
 VaSTS PRFVVIDDGWQSVNFDDE--DPNEDAKNLVVGGEQMTARLHRLHFEEDKFRKYQKGLLGNAPSFNPKTI 269
 GmSTS19.1 PRFVIIDDGWQSVNFDGD--DPNVDAKNLVVGGEQMTARLHRLHFEEDKFRKYQKGLLGNAPSFNPKTI 269
 SaSTS PRFLIIDDGWQSI INIDGQ--DPNEDAKNLVVGGEQMTARLHRLHFEEDKFRKYQKGLLGNAPSFNPKTI 322
 AmSTS PRFLIIDDGWQSI NNDDN--DPNEDAKNLVVGGEQMTARLHRLHFEEDKFRKYQKGLLGNAPSFNPKTI 328
 AtSTS PKFVIIDDGWQSI NFDGD--ELDKDAENLVVGGEQMTARLHRLHFEEDKFRKYQKGLLGNAPSFNPKTI 333

PsRFS -----DNGG 324
GmRFS06 -----KDS- 297
PrRFS -----KVPS 290
GmRFS05.1 -----REG- 298
GmRFS05.2 -----E 263
PsSTS1 **TDLILKGIIEHEKLRKKREEAISKSSD--LAEIESKIKKVVKEIDDLFGGE-----QFSSGEKSEMKS** 380
MtSTS **KELITKGIIEHEHLERQREEAVLSKSSD--LAEIESKIKQLVKEIDDLFGGE-----ESSGAPKSELK-** 377
VaSTS **KELISKGIEAEHLGKQ-AAAISAGGSD--LAEIETLMVVKVREIDDLFGGKG--KESNE-SGGCCCKAAE** 383
GmSTS19.1 **KELIAKGIIEVERLGLRDEAVSSGVSDLSTEIESRIVKVKKEIDDLFGGEG--KENKELCGGCCCKANE** 387
SaSTS **KLLISKAIEIEGVEKARDKAIQSGITD--LSQYEIKLKKLNKELDEMFGGGGNDKESGSKGSCSDCSCKSQ** 390
AmSTS **KLLISKAIEIEVAEKARDKAAQSGVTD--LARYEAEIEKLTKELDQMFGGGG--EETSSGKSCSSCSCKSD** 395
AtSTS **KMLIYKATERIQAIILRRKLVKESGEQ-DLTELDEKIKILSELNAMFDEVE-----KEESLGSDDVS** 395

PsRFS KKGLVGFVRDLKEEFRSVESVYVWHALCGYWGVRPKVCG-MPEAKVVVPKLSPGKMTMEDLAVDKIVE 393
GmRFS06 EKGMGAFVRDLKEQFRSVEQVYVWHALCGYWGVRPKVPG-MPQAKVVTPKLSNGLKTMKDLAVDKIVS 366
PrRFS GRGMSAFIRDLKEEFGTIEHVYIWHAVCGYWGVRPAVGGNMPESRVISPKLSPSLQMTMEDLAVDKIVN 360
GmRFS05.1 NKGMGAFVRELKEEFGSLEYVYVWHAFVCGYWGVRPKVPG-MPEATVVPTKLSPGAEMTMTDQAVVKIME 367
GmRFS05.2 GKGLKGFVRELKEEFGSVEYVYVWHALCGYWGVRPVGAV-MAEAAVEKPKLTGLKGTMEDLAVDKIVN 332
PsSTS1 EYGLKAFTRDLRTRFKGLDDVYVWHALCGAWGVRPE-TTHLDTKIVP-CKLSPGLDGTMEDLAVVEISK 448
MtSTS DYGLKAFTRDLRTRFKGLDDVYVWHALCGAWGVRPG-ATHLDAKIIIP-CKLSPGLDGTMHDLAVVNIK 445
VaSTS CGGMKDFTTDLRTRTEFKGLDDVYVWHALCGGWGVRPG-TTHLDSKIIIP-CKLSPGLVGTMKDLAVDKIVE 451
GmSTS19.1 CGGIKAFIRDLRTRTEFKGLDDVYVWHALCGSWGVRPG-ATHLNSKITP-CKLSPGLDGTMQDLAVVKIVE 455
SaSTS NSGMKAFNTDLRTRTFKGLDDIYVWHALAGAWGVRPG-ATHLNAKIEP-CKLSPGLDGTMTDLAVVKILE 458
AmSTS NFGMKAFTRDLRTRTFKGLDDIYVWHALAGAWGVRPG-ATHLNAKIVP-TNLSPLDGTMTDLAVVKIIE 463
AtSTS GSGMAAFTRDLRTRFKSLDDIYVWHALCGAWGVRPE-TMMDLAKAVAP-FELSPSLGATMADLAVDKIVE 464

PsRFS NGVGLVPPNLAQEMFDGIHSHLESAGIDGVKVDVIHLELLESEYGGRVELAKAYYKALTSSVNKHFKGN 463
GmRFS06 NGVGLVPPHLLHLLYELHSHLESAGIDGVKVDVIHLELLESEYGGRVELAKAYYKALTASVKKHFKGN 436
PrRFS NGVGLVQPELAYKMYEGLHSHLESAGIDGVKVDVIHLELLESEYGGRVALAEAYYKALTASVRKHFKGN 430
GmRFS05.1 IGVGLVPPHRAHELYEGLHSHLESVGDGVKVDVIHLELLESEYGGRVELAKAYYKALTASVRKHFKGN 437
GmRFS05.2 NGVGVPPELVGEMYERLHAHLESAGIDGVKVDVIHLELLESEYGGRVDMAKAYYKALTASVRKHFKGN 402
PsSTS1 ASLGLVHPHQANELYDSMHSYLAESGITGVKVDVIHLSLEYVCDEYGGRVDLAKAVYEGGLTKSIVKNFNGN 518
MtSTS GAIIGLVHPDHVDFYDSMHSFLAESGVTGVKVDVIHTLKYVCDEYGGRVNLAKAYYEGGLTKSIKNFNGS 515
VaSTS GSIIGLVHPHQANDLYDSMHSYLAQGTGVKVDVIHLSLEYVCEYGGRVELAKAYYDGLTNSIIKNFNGS 521
GmSTS19.1 GSIIGLVHPDQANDLYDSMHSYLAQSGVTGVKVDVHLSLEYVCEYGGRVELAKAYYDGLTNSIIKNFNGS 525
SaSTS GSIIGLVHPDQAEFDYDSMHSYLSKVGITGVKVDVIHTLEYVSENYGGRVELGKAYYGLSKLKKNFNGS 528
AmSTS GSTGLVDPDQAEFDYDSMHSYLSVSGITGVKVDVIHTLEYISEDYGGRVELAKAYYGLSKSLAKNFNGT 533
AtSTS AGIIGLVHPSKAHEFDYDSMHSYLSVSGITGVKVDVIFQTLSESLAEHGGRVELAKAYYDGLTESMIKNFNGT 534

PsRFS GVIASMEHCNDFLLGTEAIALGRVGDDFWCCDPSGDPNGTYWLQGCHMVHCAYNLSLWGMNFIHPDWD MF 533
GmRFS06 GVIASMEHCNDFLLGTEAIALGRVGDDFWCTDPSGDPNGTYWLQGCHMVHCAYNLSLWGMNFIQPDWDMF 506
PrRFS GVIASMEHCNDFMFLGTEAIALGRVGDDFWCTDPSGDPNGTYWLQGCHMVHCAYNLSLWGMNFIHPDWD MF 500
GmRFS05.1 GVISSMQQCNDFFLGTETISLGRVGDDFWCTDPAGDPNGTYWLQGCHMVHCAYNLSLWGMNFIHPDWD MF 507
GmRFS05.2 GVIASMEHCNDFMLLGTETISLGRVGDDFWCTDPYGDPNPTFWLQGCHMVHCAYNLSLWGMNFIHPDWD MF 472
PsSTS1 GMIASMQHCNDFFLGTETISLGRVGDDFWQDPNGDPMGSFWLQGVHMHCSYNSLWGMQMIQPDWDMF 588
MtSTS GI IASMQQCNDFFLGTQVSMGRVGDDFWQDPNGDSMGVFWLQGVHMHCSYNSLWGMQMIQPDWDMF 585
VaSTS GI IASMQQCNDFFLGTQIPFGRVGDDFWQDPNGDPMGVFWLQGVHMHCSYNSLWGMQMIQPDWDMF 591
GmSTS19.1 GI IASMQQCNDFFLGTQIPMGRVGDDFWQDPNGDPMGVFWLQGVHMHCAYNLSLWGMQMIQPDWDMF 595
SaSTS GLISSMQQCNDFFLLGTEQISLGRVGDDFWQDPNGDPMGVFWLQGVHMHCAYNLSLWGMQMIQPDWDMF 598
AmSTS GLISSMQQCNDFFLLGTEQISLGRVGDDFWQDPNGDPMGVFWLQGVHMHCAYNLSLWGMQMIQPDWDMF 603
AtSTS DVIASMQQCNEFFLATKQISIGRVGDDFWQDPYGDYVWQGVHMHCSYNSIIMGMQMIQPDWDMF 604

PsRFS **QSTHPCAEFHAASRAISGGPVYVSDCVG--NHNFKLLKSFVLPDGSILRQHYALPTRDCLFEDPLHDGK** 601
GmRFS06 **QSTHPCAEFHAASRAISGGPVYVSDCVG--KHNFKLLKSLALPDGTILRQHYALPTRDCLFEDPLHDGK** 574
PrRFS **QSTHPCAEFHAASRAISGGPIYVSDSVG--KHNFKLLKALVLPDGSILRQHYALPARDCLFEDPLHDGK** 568
GmRFS05.1 **QSDHACAEFHAASRAISGGPIYVSDSVG--KHNFKLLKLVLPDGSILRQHYALPTRDCLFVDPDLHDGK** 575
GmRFS05.2 **QSTHPCAFAHAASRAISGGPIYISDITVG--NHNFELKLTALPDGSILRQHYALPTRDCLFADPLHDGK** 540
PsSTS1 **QSDHCAKFAHAGSRAICGGPIYVSDNVG--SHDFDLIKKLVFPDGTIPKCIYFPLPTRDCLFKNPLFDHT** 656
MtSTS **QSDHICAKFAHAGSRAICGGPVYVSDNVG--SHDFDLIKKLVFPDGTIPKCIHFPLPTRDCLFKNPLFDQT** 653
VaSTS **QSDHECAKFAHAGSRAICGGPVYVSDSVG--SHDFDLIKKLVFPDGTVPKCIYFPLPTRDCLFRNPLFDQK** 659
GmSTS19.1 **QSDHCAKFAHAGSRAICGGPVYVSDSVG--SHDFDLIKMLVFPDGTVPKCIHFALPTRDCLFKNPLFDQK** 663
SaSTS **QSDHCSAKFAHAGSRAICGGPVYVSDSLG--GHDFDLKLVFPDGTIPKCIHFALPTRDCLFKNPLFDSK** 666
AmSTS **QSDHPGGYFAHAGSRAICGGPVYVSDSLG--GHDFDLKLVFPDGTIPKCIHFALPTRDCLFKNPLFDSK** 671
AtSTS **QSDHCAEYHAASRAICGGPVYVSDSLHGLKASHNFDLIIKLAFFDGTIPRCVHYALPTRDCLFKNPLFDKE** 674

PsRFS TMLKIWNLNKYAGVLGFLNCGGGWCPVTRRNKASAEFSHAVTICYASPEDIEWCNGKTPMD--IKGVDFV 669
GmRFS06 TMLKIWNLNKYTGVLGFLNCGGGWCPVTRRNKASAEFSQVTCTCLASPQDIEWNSGKSPIC--IKGMNVF 642
PrRFS TMLKIWNLNKYTGVLGIFNCGGGWCPVARRNKSANQFSQVSTCSASPKDIEWNSGKSPIS--VKGVDVF 636
GmRFS05.1 TMLKIWNLNKCSGVLGFLNCGGGWCPVTRRNKSSSDYSHSVTCFASPQDIEWGKGHFVC--IKGVDFV 643
GmRFS05.2 TMLKIWNLNKYTGVLGFLNCGGGWFREIRSNKCAAEFSHRVSTKTNIKDIEWDSGKNPIS--IEGVQLF 608
PsSTS1 TVLKIWNFNKYGGVIGAFNCGAGWDPIMQKFRGFPECYKPIPGTVHVTEVEWDQKETS--HLGKAEY 724
MtSTS SVLKIWNFNKYGGVIGAFNCGAGWDPKEHKFRGFPECYKPIPGTVHVTEVEWDQKETS--AEEY 707
VaSTS TVLKIWNFNKYGGVIGAFNCGAGWDPKGGKFRGFPECYKAIISCTVHVTEVEWDQKKEAE--HMGKAEY 727
GmSTS19.1 TVLKIWNFNKYGGVIGAFNCGAGWDPKMKKIKGFSECYRPIISCTVHVTEVEWDQKKEAV--HMGKAEY 731
SaSTS TILKIWNFNKYGGVIGAFNCGAGWDPKEQRIKGYSECYKPLSGSVHVSIDIEWDQKVEAT--KMGKAEY 734
AmSTS TILKIWNFNKYGGVIGAFNCGAGWDPKEQRIKGYSCYKPLSGSVHVSIDIEWDQKVEAS--EMGEAEY 739
AtSTS SILKIWNFNKYGGVIGAFNCGAGWSPPEHRFKGYKCYTTSVGTVHVSDIEWDQNPAAQSQVTYTGDY 744

PtRFS	AVYFFKEKKLSLMKCS-D-RLEVSLEFFSFELMTVSP	PLKVF	SKRLIQFAP	I	GLV	NMLNSGGAV	QSLEF	DDS	738																												
GmRFS06	AVYLFKDHKLKLMKASE-KLEVSLEPFTFELLTVSP	VIVL	SKKLIQFAP	I	GLV	NMLNTGGAI	QSM	EFDNH	711																												
PtRFS	AVYMFKEKKVRLKSS-E-KLEISLEPFNYDLLTVSP	VT	LPKRSIQFAP	I	GLV	NMLNTGGAI	QSM	MVDD	705																												
GmRFS05.1	AVYMFKDDKLLKLYTE-SVEVSLEPFSCCELLTVSP	VVIL	LPKRSIQFAP	I	GLV	NMLNSGG	SIMS	LEFDQ	712																												
GmRFS05.2	ASYFSQAKKLILSAPSDDSEEISLEPFNFELITVSP	VT	LPKRSIQFAP	I	GLV	NMLNTGGAV	QSLAF	DEG	678																												
PtSTS1	VVYLNQAEELSLMTLKSEPIQFTIQPSTFELYSFV	VP	TKLGG- IKFAP	I	GLT	NMFNSGGT	VIDLE	YV--	791																												
MtSTS	VVYFNQAEELRLMTPKSEPIKYYIIQPSTFELYNF	VPL	TKLGGN- IKFAP	I	GLT	NMFNSGGT	VLDLE	YA--	774																												
VaSTS	VVYLNQAEVLHLMTPVSEPLQFTIQPSTFELYNF	VVP	VEKLGSSNIK FAP	I	GLT	NMFNSGGT	IQE	LEYI--	795																												
GmSTS19.1	VVYLNQAEELHFMTPKSEPLQFTIQPSTFEIYNF	VVP	VEKLGGS- IKFAP	I	GLT	NMFNSGGT	IQE	LECV--	798																												
SaSTS	AVYLTESEKLLLTTPESDPIPFTLKSTTFEIFS	FVP	IKKLGQG- VKFAP	I	GLT	NLFNSGGT	IQ	GVVYD--	801																												
AmSTS	AVYLSEAEKLSLATRSDPIKITIQSSTFEIFS	FVP	IKKLGEG- VKFAP	I	GLT	NLFNAGGT	IQ	GLVYN--	806																												
AtSTS	LVIYKQQSEELFMNSKSEAMKITLEPSAFD	LLS	FVPTVTELV	SSG	VR	FAP	LGLIN	MFCV	GTQ	DMK	V	TG-	813																								
PtRFS	ASLVKIGVRGCGELSVFASEKPVCKIDGVSVEFDYEDK--	MVR	QILWPG-SSTLS	LVE	FLF	798																															
GmRFS06	IDVVKIGVRGCGEMKVFASEKEVSCKLDG	VVVK	FDYEDK--MLRV	QVP	WPS-ASK	LSM	VEFLF	771																													
PtRFS	ESLIRIGVKGSGEMRVFASGNPVSKIDGVD	VEFC	FHDQ--MVTI	QVP	WPS-SP	KL	SVM	FLF	765																												
GmRFS05.1	ENLARIGVRGHGEMRVFASEKPEVSKIDGES	VEFDY	VDR--TVRL	QV	SWPC-SS	RLS	VVE	YLF	772																												
GmRFS05.2	QNLVEVGLRGTGEMRVYASEKPRTRCIDGKE	VD	FEYEGS--MVNI	QVP	WPG-SS	KL	STV	QYVF	738																												
PtSTS1	GNGAKIKVKGGSFLAYSSSEPKKFQLNG	CEVD	FEWLGD-GKLC	VNP	WIE	EAC	GV	S	D	M	E	I	F	853																							
MtSTS	ESGAKIQVKGGSFLAYSSSEPKKFQLNG	SEVAF	EWLGD-GKLS	LNV	SWIE	E	AS	GV	S	E	L	A	I	F	836																						
VaSTS	EKDVKVVKGGGRFLAYSTQSPKKFQLNG	S	DAAFQWLDP-GKLT	LNL	AWIE	E	ND	GV	S	D	L	A	I	F	857																						
GmSTS19.1	EKGAKVVKGDGRFLAYSSSEPKKFQLNG	S	DVAFEWLDP-GKLT	LNL	AWIE	E	NG	GV	S	D	L	A	I	F	860																						
SaSTS	EGVAKIEVKGDGKFLAYSSSVPKRSYLN	GE	EVEYKWSGN-GK	VE	D	V	W	Y	E	E	C	G	I	S	N	I	T	F	V	863																	
AmSTS	EGIAKIEVKGDGKFLAYSSSVPKKAYV	NG	AEKVF	AW	S	G	N	-GK	L	E	L	D	I	T	W	Y	E	E	C	G	I	S	N	V	T	F	V	868									
AtSTS	DNSIRVDVKGGRFMAYSSSAFVVKCYL	N	D	K	E	A	E	F	K	W	E	E	T	G	K	L	S	F	F	V	P	W	Y	E	E	S	G	G	I	S	H	L	S	F	T	F	876

Supplementary Fig. 1 Alignment of the sequences of predicted RFSs and STSs encoded in the soybean genome with *P. trichocarpa* (XP_002321625), *P. sativum* (CAD20127; CAC38094), *M. truncatula* (Medtr4g136550), *V. angularis* (CAB64363), *S. affinis* (CAC86963), *A. meridionalis* (CAD31704) and *Arabidopsis* (BAF00996). Residues marked in gray are identical for all sequences aligned; motifs conserved are highlighted in black, and the diagnostic sequences for STS are in bold, corresponding the positions 305 to 383 of GmSTS19.1

CAPÍTULO 2

CONSTRUÇÃO DE CASSETES DE EXPRESSÃO PARA SILENCIAMENTO DO GENE DA ESTAQUIOSE SINTASE DE SOJA (*Glycine max* L. Merrill), VIA INTERFERÊNCIA POR RNA

RESUMO

A soja é a uma excelente fonte de proteínas, óleos poliinsaturados e fibras, no entanto, existem fatores que limitam o seu consumo na alimentação, tais como os oligossacarídeos de rafinose (ROs), responsáveis por distúrbios gastrointestinais. O silenciamento de genes específicos da via de biossíntese de ROs, através de interferência por RNA, pode resultar em variedades de soja com reduzido conteúdo de ROs. O objetivo deste trabalho foi construir cassetes de expressão, visando o silenciamento gênico da enzima estaquiose sintase (STS) em soja, via interferência por RNA. Após extração do RNA total de sementes de soja, o cDNA foi sintetizado por RT-PCR. O gene *GmSTS*, 2.6 kb, foi amplificado por PCR utilizando-se oligonucleotídeos iniciadores específicos, clonado e sequenciado. Fragmentos internos, correspondendo aos fragmentos *sense* e *antisense*, cujas sequências correspondem a duas regiões distintas do gene estaquiose sintase apresentando alta e baixa identidade com genes de rafinose sintases de soja, de 366 pb e 387pb, respectivamente, foram amplificados utilizando-se oligonucleotídeos iniciadores internos específicos. Estes fragmentos foram clonados no vetor pGEM-T *Easy* e sequenciados. Em seguida, foram clonados nos vetores de expressão pKANNIBAL e pBKN flanqueados pelo promotor constitutivo 35SCaMV e pelo promotor semente específico da subunidade α da proteína β -conglucina, respectivamente, e pela região terminadora OCS, e espaçados pelo

íntron PDK. Os quatro cassetes obtidos foram transferidos para o vetor pCAMBIA 3301, para inserção no genoma de soja, em uma próxima etapa, via *Agrobacterium tumefaciens*.

1. INTRODUÇÃO

A soja (*Glycine max* L. Merrill) é uma das culturas economicamente mais importantes no mundo, sendo considerada uma fonte global de proteínas e óleo (Schmutz et al., 2010). Entretanto, os fatores antinutricionais limitam o uso desta leguminosa e de seus derivados na alimentação humana e animal. Oligossacarídeos de rafinose (ROs), principalmente rafinose e estaquiose, são considerados fatores antinutricionais porque não podem ser digeridos pelo trato gastrointestinal humano e de animais monogástricos devido a ausência da enzima α -galactosidase no intestino delgado, necessária para hidrólise das ligações α -1,6 presentes nos ROs (Kumar et al., 2010). A remoção de rafinose e estaquiose tem demonstrado aumentar significativamente o valor nutricional de produtos alimentares derivados de soja (Leske et al., 1993). O efeito da redução de rafinose e estaquiose em dietas tem sido associado ao aumento da energia metabolizável e a redução da produção de flatulência em humanos (Suarez et al., 1999), frangos (Parsons et al., 2000), cachorros (Zuo et al., 1996) e em porcos (Smiricky et al., 2002).

A redução de ROs em soja, através do melhoramento genético, tem recebido considerável atenção nos últimos anos. Entretanto a maioria das populações melhoradas de soja tem sido construída a partir do mutante PI200508, no qual o baixo conteúdo de rafinose e estaquiose foi associado a uma mutação em um gene que codifica para rafinose sintase (Dierking e Bilyeu, 2008). Ainda não foram descritas, em nosso conhecimento, variedades de soja cujo baixo conteúdo de ROs esteja associado a alterações na expressão do gene ou na atividade da estaquiose sintase. Considerando que estaquiose é o ROs mais abundante nas sementes de soja e que tem demonstrado não ser requerido para a eficiente germinação das mesmas (Dierking e Bilyeu, 2009), o silenciamento gênico da estaquiose se torna uma alternativa viável para reduzir o

conteúdo de ROs, sem alterar o teor dos demais carboidratos, nas sementes desta leguminosa.

Uma ferramenta molecular que tem demonstrado ser altamente eficiente para silenciamento de genes específicos é a interferência por RNA ou *RNA interference* (RNAi). As aplicações de RNAi nas plantas estão sendo direcionadas para melhoria da produtividade e/ou valor nutricional em várias espécies vegetais (Liu et al., 2002; Kusaba et al., 2003; Davuluri et al., 2005; Sunilkumar et al., 2006; Houmard et al., 2007). Em soja, RNAi tem sido aplicado para produção de sementes contendo óleo com maior estabilidade em altas temperaturas (Wang e Xu, 2008; Flores et al., 2008) e com reduzido conteúdo de fitato (Nunes et al., 2006); não tendo sido ainda relatado o uso desta tecnologia para a redução do conteúdo de ROs nesta leguminosa.

O objetivo deste trabalho foi isolar fragmentos do gene que codifica a enzima estaquiose sintase em soja (*GmSTS*), em duas regiões distintas com alta e baixa identidade de seqüência com genes da rafinose sintase em soja, para a construção de cassetes de silenciamento gênico por RNA de interferência.

2. MATERIAL E MÉTODOS

2.1. Material Vegetal

Sementes de soja (*Glycine max* L. Merrill), da variedade comercial CAC-1, foram usadas para este estudo. As plantas foram crescidas sob aquecimento controlado e fotoperíodo de 14 horas de luz, em casa de vegetação da Universidade Federal de Viçosa, Viçosa – Minas Gerais.

As sementes foram coletadas em um período de aproximadamente três meses (plantio escalonado), separadas em oito estágios de desenvolvimento, que compreenderam todo o período de enchimento do grão, congeladas em nitrogênio líquido e armazenadas a -80°C. Estes estágios foram determinados com base no peso de matéria fresca da semente: 1° - 0 a 75 mg; 2° - 76 a 150 mg; 3° - 151 a 225 mg; 4° - 226 a 300 mg; 5° - 301 a 375 mg; 6° - 376 a 450 mg; 7° - 451 a 525 mg e 8° - representado por sementes maduras.

2.2. Extração de RNA total

Baseado no RT-PCR semi-quantitativo realizado previamente para análise da expressão gênica (Fialho, 2007), a extração de RNA total das sementes foi conduzida a partir de sementes do 7° estágio, no qual foi observado maior acúmulo de transcrito. Esta extração foi conduzida de acordo com Sambrook et al. (1989), adaptando o protocolo original, sendo todas as etapas realizadas a 4°C em condições livres de RNAses.

Para obtenção do RNA total das sementes de soja, cerca de 4 g de sementes foram maceradas em nitrogênio líquido. Este macerado foi homogeneizado com 18 mL

de tampão NTES (NaCl 100 mM; Tris-HCl 10 mM, pH 7,5; EDTA 1 mM e SDS 1%), 6 mL de fenol e 6 mL de clorofórmio:álcool isoamílico (24:1). Os tubos foram vigorosamente agitados em agitador vórtex por 15 min e a fase aquosa foi separada da fase orgânica depois de centrifugados a 4°C por 10 min a 8.000 x g. À fase aquosa foi adicionado 0,1 volumes de acetato de sódio 2 M e 2,5 volumes de etanol 96% e, em seguida, a mistura foi incubada por 1 hora a -20°C. Foi realizada uma nova centrifugação a 8.000 x g por 15 min e o precipitado foi lavado com etanol 70%, ressuspenso em 2,5 mL de água DEPC (água deionizada tratada com dietil pirocarbonato), e novamente centrifugado por 5 min a 5.000 x g, a 4°C. O sobrenadante foi transferido para novo tubo onde foram adicionados 2,5 mL de cloreto de lítio 4M. O tubo foi incubado por cerca de 12 horas a 4°C para promover a precipitação do RNA. As amostras foram então centrifugadas a 8.000 x g por 30 min e o precipitado foi ressuspenso em 1,8 mL de água DEPC e acrescido de 0,2 mL de acetato de sódio 2M e 3,6 mL de etanol 96%. Após precipitação por 6 horas a -20°C, a solução foi centrifugada a 8.000 x g por 10 min e em seguida, o precipitado foi lavado com etanol 70% e depois seco e ressuspenso em água DEPC.

O RNA extraído foi quantificado em espectrofotômetro *DU 650 BECKMAN*, a 260 nm, sendo a concentração do RNA expressa em $\mu\text{g}.\mu\text{L}^{-1}$. A integridade do RNA total foi determinada através de eletroforese em gel de agarose 1,5% em tampão TBE (Tris-borato 90 mM e EDTA 1 mM, pH 8,0) contendo $0,1 \mu\text{g}.\text{mL}^{-1}$ de brometo de etídeo. O padrão de bandas do RNA foi visualizado sob luz ultravioleta e fotografado com o sistema de fotodocumentação.

As amostras quantificadas foram tratadas com *RQI DNase RNase-free* (PROMEGA), conforme as recomendações do fabricante. Sendo assim, as amostras foram incubadas em tampão de reação (Tris-HCl 40 mM, pH 8,0; MgSO_4 10 mM e CaCl_2 1 mM), a 37°C, por 45 min, e extraídas com igual volume de fenol e de clorofórmio:álcool isoamílico após centrifugação a 11.000 x g, a 4°C, por 2 min. Após transferência da fase aquosa para um novo tubo, foi realizada nova lavagem com clorofórmio:álcool isoamílico (24:1). A fase aquosa foi submetida à precipitação com acetato de sódio 3M e etanol 96%, a -20°C, por 1 hora, e depois centrifugada a 12.000 x g por 15 min, a 4°C. O precipitado foi lavado com etanol 70%, seco e ressuspenso em água DEPC. As amostras de RNA foram devidamente quantificadas no espectrofotômetro a 260 nm.

Para síntese da primeira fita de cDNA foi utilizado o *Kit SuperScript™ First-Strand Synthesis System* (INVITROGEN) segundo as recomendações do fabricante. O RNA total (2 µg), usado como molde, foi incubado por 20 min, a 70°C com 1 µL de oligonucleotídeos Oligo(dT)12-18. Para reação foram adicionados tampão de PCR (Tris-HCl 20 mM, pH 8,4 e KCl 50 mM), MgCl₂ mM, dNTP's 0,5 mM e DTT 5mM. Esta reação foi aquecida e mantida a 42°C por 5 min. Em seguida a reação foi transferida para o gelo e foram adicionadas 200U da enzima transcriptase reversa e as amostras foram incubadas a 42°C por 50 min e a 70°C por 15 min para síntese da primeira fita de cDNA. Posteriormente, o RNA foi degradado pela adição de 1U de RNase H sob incubação a 37°C por 20 min. Como controle negativo, para cada amostra foram realizadas reações que continham todos os reagentes, exceto a enzima transcriptase reversa.

2.3. Construção de oligonucleotídeos iniciadores específicos e isolamento dos fragmentos diferentes do gene que codifica para estaquiase sintase para a construção do cassete de silenciamento

A escolha dos fragmentos do cDNA do gene que codifica a enzima estaquiase sintase em soja a ser amplificada foi determinada com base em análise *in silico* através do programa TAGG (<http://workbench.sdsc.edu>) e BLOCK-iT™ RNAi Designer (INVITROGEN) (<https://rnaidesigner.invitrogen.com/rnaiexpress/>). O mapa de restrição das sequências foi construído com auxílio do programa TAGC e permitiu a escolha das enzimas de restrição adequadas para os experimentos de clonagem. O programa BLOCK-iT™ RNAi Designer possibilitou identificar os siRNAs candidatos que poderiam se formar a partir de cada seqüência de interesse em uma ordem de classificação. Dessa maneira, as regiões escolhidas para amplificação estão associadas aos melhores siRNAs obtidos.

Para construção dos oligonucleotídeos iniciadores específicos utilizou-se como base a seqüência nucleotídica de um fragmento de cDNA correspondente ao gene da estaquiase sintase de soja previamente descrito (FJ968750). Estes oligonucleotídeos específicos foram desenhados a partir de duas regiões distintas do gene *GmSTS*: uma região que apresenta alta identidade de seqüência e uma região que apresenta baixa identidade de seqüência com genes rafinose sintases de soja. O programa utilizado para a construção dos oligonucleotídeos foi o *Primer3 Input Program*:

(<http://frodo.wi.mit.edu/primer3/>) (Rozen e Skaletsky, 2000). Sítios de restrição reconhecidos pelas enzimas *Xho* I - *Kpn* I e *Xba* I - *Cla* I foram adicionados em seus terminais opostos, para amplificação dos fragmentos *sense* e *antisense*, respectivamente. As seqüências dos oligonucleotídeos estão descritas na Tabela 1.

Tabela 1: Seqüência dos oligonucleotídeos iniciadores construídos para amplificação das diferentes regiões do gene que codifica para estaquiase sintase em soja para obtenção dos fragmentos *sense* e *antisense*.

Gene alvo	Fragmentos	Oligonucleotídeos Iniciadores	Seqüência	Sítio de Restrição
<i>GmSTS</i>	<i>Sense*</i>	ESTsense-F	5'-GAA <u>TCT CGA GCG</u> AGC TGG CAA AGG CTT ATT-3'	<i>Xho</i> I
		ESTsense-R	5'-GGC <u>GGT ACC</u> TCA TGA GAG CCC ACA CTG TC-3'	<i>Kpn</i> I
	<i>Antisense*</i>	ESTantisense-F	5'-GAA <u>TTC TAG ACG</u> AGC TTG CAA AGG CTT ATT-3'	<i>Xba</i> I
		ESTantisense-R	5'-GGC <u>ATC GAT</u> TCA TGA GAG CCC ACA CTG TC-3'	<i>Cla</i> I
<i>GmSTS</i>	<i>Sense**</i>	STSense-F	5'-GGC <u>CTC GAG</u> TGC CAA AAA TTG CTG ACA AG-3'	<i>Xho</i> I
		STSense-R	5'-GGC <u>GGT ACC</u> CAG ACG AAA CAG CCT CAT CA-3'	<i>Kpn</i> I
	<i>Antisense**</i>	STSantisense-F	5'-GGC <u>ICT AGA</u> TGC CAA AAA TTG CTG ACA AG-3'	<i>Xba</i> I
		STSantisense-R	5'-GGC <u>ATC GAT</u> CAG ACG AAA CAG CCT CAT CA-3'	<i>Cla</i> I

*Região gênica com alta identidade de seqüência com genes rafinose sintase de soja.

** Região gênica com baixa identidade de seqüência com genes rafinose sintase de soja.

As regiões sublinhadas correspondem às seqüências de reconhecimento e clivagem das enzimas de restrição.

2.4. Amplificação dos fragmentos gênicos e clonagem no vetor pGEM T-Easy

As reações de PCR foram realizadas, cada uma contendo uma combinação dos oligonucleotídeos iniciadores, com o objetivo de amplificar os fragmentos *sense* e *antisense* nas diferentes regiões de interesse do gene que codifica para a enzima estaquiase sintase. Para aperfeiçoar as condições da PCR, as temperaturas de anelamento dos oligonucleotídeos iniciadores foram testadas.

As reações de amplificação dos fragmentos *sense* e *antisense* correspondentes a região gênica com alta identidade de seqüência com os genes da rafinose sintase de soja, foram realizadas a partir do cDNA de sementes de soja e conduzidas no termociclador *MasterCycle Gradient* (EPPENDORF, Hamburg, Alemanha) com período inicial de desnaturação de 94°C por 2 min, seguido por 35 ciclos (94°C por 30 s; 50°C por 30 s; 72°C por 40s) e um período adicional de extensão a 72°C por 4 min. Cada reação, com volume final de 25 µL, continha 1,0 µL de cDNA como molde, 2,5 µL de tampão IB 10X (Tris pH 8,4 100 mM; KCl 500 mM; Triton X-100 1%; MgCl₂ 15 mM), 0,2 µM de oligonucleotídeos iniciadores, dNTP's 0,2 mM, 1U de *Taq* DNA Polimerase

(PHONEUTRIA). Os fragmentos amplificados foram purificados através do *kit Wizard® SV Gel PCR Clean-Up System* (PROMEGA). As reações de amplificação dos fragmentos *sense* e *antisense* a partir de uma seqüência do gene *GmSTS* com baixa identidade com genes rafinose sintase foram conduzidas de forma similar, entretanto a temperatura de anelamento dos oligonucleotídeos iniciadores utilizada foi de 60°C.

2.5. Transformação de *Escherichia coli* por choque térmico e diagnóstico molecular dos transformantes

Após amplificação e purificação, os fragmentos *sense* e *antisense* foram ligados ao vetor pGEM T-*Easy* através de uma reação de ligação usando a enzima T4 DNA ligase contida no *kit pGEM® T-Easy Vector System* (PROMEGA), de acordo com as recomendações do fabricante. As células de *E. coli* DH5 α ultracompetentes foram transformadas por choque térmico de acordo com Sambrook et al. (1989). Foram adicionados 5 μ L de reação de ligação à 100 μ L de células ultracompetentes e esta mistura foi incubado no gelo por 30 min. Após choque térmico, a 42°C por 50 s, as células permaneceram no gelo por mais 2 min e, em seguida, foram adicionados 800 μ L de meio LB (triptona 10 g.L⁻¹; extrato de levedura 5 g.L⁻¹; NaCl 10 g.L⁻¹). Após incubação por 1 h a temperatura de 37°C a 150 rpm, as células foram concentradas por centrifugação a 1.500 x g por 10 min, ressuspensas em meio LB, e plaqueados em meio LB com ampicilina 50 μ g. μ L⁻¹, 1 mg de X-Gal e 10 mM de IPTG. As placas foram incubadas por 14 h a 37°C, para crescimento e multiplicação celular.

Para cada transformação foi realizada a seleção de 10 colônias aleatoriamente e o diagnóstico para confirmação de transformantes foi feito por meio de PCR seguindo as condições anteriormente descritas. Foi realizada corrida eletroforética em gel de agarose 1%, em TBE, com brometo de etídeo 0,1 μ g.mL⁻¹ para análise dos fragmentos amplificados. Após eletroforese, a integridade dos fragmentos amplificados foi determinada sob luz ultravioleta e o gel foi fotografado com o sistema de fotodocumentação.

As colônias que apresentaram amplificação dos fragmentos esperados foram selecionadas e transferidas para tubos com meio LB e ampicilina 50 μ g.mL⁻¹ e incubados a 37°C e 180 rpm por 14 h. Estes clones foram estocados, posteriormente, a -80°C, em glicerol 25%. O DNA plasmidial foi extraído com o *kit Wizard® Plus SV Miniprep* (PROMEGA) segundo as recomendações do fabricante. A clonagem foi

confirmada através de reações de PCR, realizadas como anteriormente descrito, clivagem por enzimas de restrição e sequenciamento.

As clivagens dos fragmentos *sense* foram realizadas com 1.000 ng de DNA plasmidial, 5U da enzima *Kpn* I e tampão de clivagem (Tris-HCl 10 mM pH 7,5; MgCl₂ 7 mM; KCl 50 mM; DTT 1 mM) e incubadas a 37°C, em banho-maria, por 1 hora. Em seguida foram adicionados 1,0 µL de *Xho* I NaCl 3M e a reação foi novamente incubada a 37°C por 1 hora. Para clivagens dos fragmentos *antisense*, foram utilizados 1.000 ng de DNA plasmidial, tampão de clivagem (Tris-acetato 33mM pH 7,9; acetato de magnésio 10mM; acetato de potássio 66mM; BSA 0,1 g.mL⁻¹; DTT 1mM; glicerol 50%) e 5U das enzimas *Kpn* I e *Cla* I. O sequenciamento dos clones recombinantes foi realizado pela empresa Macrogen (Seoul, Coreia do Sul) e as sequências obtidas foram analisadas pelo programa computacional *Sequencher* versão 4.1.4, sendo o alinhamento com outras sequências disponíveis no *Genbank* realizado através do algoritmo *blastn* e do programa ClustalW2 (Thompson et al., 1994).

2.6. Clonagem dos fragmentos específicos em vetor pKANNIBAL

As clonagens dos fragmentos *sense* e *antisense* específicos foram realizadas utilizando o vetor de clonagem pKANNIBAL. Após sequenciamento e confirmação dos fragmentos clonados, estes foram liberados do vetor pGEM T-*Easy* e inseridos no vetor pKANNIBAL (Wesley et al., 2001), de 6 kb, gentilmente cedido pelo Dr. Francisco Aragão (EMBRAPA/CENARGEN). Este vetor contém, além do promotor 35SCaMV, a região de clonagem para a sequências *sense*, o intron do gene que codifica a piruvato desidrogenase quinase (PDK), a região de clonagem das sequências *antisense* e o terminador do gene que codifica a octopina sintase (OCS). A marca de seleção deste plasmídeo corresponde ao gene *npt*, que codifica a enzima neomicina fosfotransferase II (NPT II) cujo produto confere resistência à canamicina.

Para as reações de restrição enzimática foram utilizadas enzimas de restrição *Xho* I - *Kpn* I e *Cla* I - *Xba* I para clivagem dos fragmentos *sense* e *antisense*, respectivamente, conforme descrito no item 2.5. O vetor pKANNIBAL foi clivado com as mesmas enzimas de restrição para proporcionar os mesmos sítios de clonagem e possibilitar as reações de ligação. As reações de clivagem foram analisadas em gel de agarose 1% em TBE, com brometo de etídeo 0,1 µg.mL⁻¹.

As clonagens dos fragmentos ocorreram em duas etapas: (i) subclonagem dos fragmentos *sense* e, após a confirmação deste por reação de PCR e restrição enzimática, (ii) clonagem dos fragmentos *antisense*.

As reações de ligação foram realizadas para um volume final de 10 µL, utilizando-se a relação 5:1 de DNA (50 ng) e vetor clivado (10 ng), tampão de reação (Tris-HCl pH 7,6; MgCl₂ 10 mM; ATP 1 mM; DTT 1 mM e 25 de polietileno glicol-8000) e 1U de T4 DNA ligase (INVITROGEN). As reações foram incubadas a temperatura de 4°C, por 14 h. Em seguida foram realizadas as transformações de *E. coli* DH5α por choque térmico, conforme item 2.5. O diagnóstico dos transformantes e extração do DNA plasmidial também foram conduzidas conforme descrito no item anterior, com exceção da etapa final da transformação, na qual as células foram plaqueadas em meio LB sólido contendo o antibiótico canamicina 50 µg.mL⁻¹.

2.7. Nomenclatura dos clones contendo os fragmentos *sense* e *antisense* em vetor pKANNIBAL

Os clones contendo os fragmentos *sense* e *antisense* foram denominados com base no vetor utilizado (pKANNIBAL), no promotor (35SCaMV) e nas regiões de interesse do gene *GmSTS*. Sendo assim, a nomenclatura utilizada para determinar o clone contendo os fragmentos *sense* e *antisense* amplificados a partir da região do gene *GmSTS* com alta identidade de seqüência com genes que codificam para rafinose sintases em soja foi a seguinte: pK35S-GmSTS-A, sendo A correspondente a alta. O clone contendo fragmentos *sense* e *antisense* amplificados a partir da seqüência do gene *GmSTS* com baixa identidade de seqüência com os genes rafinose sintases de soja, foi designado de pK35S-GmSTS-B, sendo B correspondente a baixa, para diferenciação.

2.8. Transferência das construções dos vetores pKANNIBAL para os vetores pCAMBIA 3301

Após a confirmação da clonagem dos fragmentos *sense* e *antisense*, as construções (promotor 35SCaMV + fragmento *sense* + intron PDK + fragmento *antisense* + terminador OCS) foram excisadas do vetor pKANNIBAL e, posteriormente, clonadas no vetor binário pCAMBIA 3301. Este vetor possui 11 kb, contém uma região de T-DNA (5.077 pb) e apresenta o promotor 35SCaMV dirigindo a expressão do gene

gus (que codifica a β -glucuronidase), o terminador do gene que codifica a nopalina sintase (NOS), além do gene *bar*, para a seleção em plantas. Esse gene codifica a enzima fosfonitricina acetiltransferase, que confere resistência a herbicidas que contenham glufosinato de amônio como composto ativo.

O vetor pCAMBIA 3301 foi clivado com enzimas de restrição *Pst I* e *Sac I*, as quais também foram utilizadas para clivar e liberar as construções anteriormente descritas. Estas clivagens foram realizadas com 5U das enzimas *PstI* e *SacI*, aproximadamente 500 ng de DNA e tampão de clivagem (Tris-HCl 50 mM, pH 8,0; MgCl₂ 10 mM; NaCl 50 mM) a 37°C por 1 hora. As análises das reações de clivagem foram realizadas em gel de agarose 1% em TBE com brometo de etídeo 0,1 $\mu\text{g}\cdot\text{mL}^{-1}$. Os fragmentos correspondentes às construções foram excisados do gel de agarose e purificados com o kit *Wizard[®] SV Gel PCR Clean-Up System* (PROMEGA). As construções liberadas e purificadas foram utilizadas para as reações de ligação com o vetor pCAMBIA 3301 previamente clivado.

As reações de ligação foram realizadas para um volume final de 10 μL , utilizando a relação de 5:1 de DNA (50 ng) e vetor (10 ng), tampão de reação (Tris-HCl pH 7,6, MgCl₂ 10 mM, ATP 1 mM, DTT 1 mM e 25% de polietileno glicol-8000) e 1U de T4 DNA ligase (INVITROGEN). As reações foram mantidas a 4°C por 14 h, conforme descrito anteriormente no item 2.6. Em seguida, foram realizadas as transformações de células de *E. coli*, por choque térmico, o diagnóstico molecular transformantes e a extração do DNA plasmidial, conforme descrito no item 2.5., excetuando a etapa final da transformação, na qual as células foram plaqueadas em meio sólido contendo canamicina 50 $\mu\text{g}\cdot\text{mL}^{-1}$. A confirmação da clonagem foi realizada por meio de PCR e reações de restrição enzimática.

2.9. Nomenclatura dos clones com as construções em vetor pCAMBIA 3301

Os clones contendo as construções foram denominados de acordo com o vetor usado (pCAMBIA 3301), o promotor (35SCaMV) e as regiões de interesse do gene *GmSTS*. Assim, as nomenclaturas utilizadas para designar as duas construções foram: pCB35S-GmSTS-A, para a construção contendo fragmentos com alta identidade de seqüência com genes rafinose sintases de soja, e pCB35S-GmSTS-B, para a construção que possui fragmentos com baixa identidade de seqüência com genes das rafinose sintases de soja.

2.10. Clonagem dos fragmentos específicos em vetor pBKN

Os fragmentos *sense* e *antisense* inseridos no vetor pGEM T-*Easy* também foram clonados no vetor pBKN, o qual foi previamente construído em nosso laboratório (Barros, 2006) pela substituição do promotor 35SCaMV, no vetor pKANNIBAL, pelo promotor semente-específico do gene que codifica a subunidade α da proteína β -conglucininina.

Para liberação dos fragmentos *sense* e *antisense* do vetor pGEM T-*Easy* e posterior clonagem no vetor pBKN foram usadas as enzimas de restrição *Kpn* I - *Xho* I e *Cla* I - *Xba* I, respectivamente, conforme descrito no item 2.5. O vetor pBKN foi clivado com as mesmas enzimas de restrição, sendo realizada primeiro a clonagem dos fragmentos *sense* e posteriormente as dos fragmentos *antisense*. Os procedimentos de clonagem, reação de ligação e diagnóstico dos transformantes foram conduzidos de acordo com o item 2.6.

2.11. Nomenclatura dos clones contendo as construções dirigidas pelo promotor do gene da subunidade α da β -conclucininina em vetores pBKN

Os clones contendo as construções dirigidas pelo promotor do gene da subunidade α da β -conglucininina, no pBKN, também foram denominadas conforme o vetor utilizado (pBKN), o promotor abreviado (β) e as regiões de interesse do gene *GmSTS*. Desta forma, a construção contendo fragmentos com alta identidade de seqüência com genes rafinose sintases de soja foi designada de pBK β -*GmSTS*-A, enquanto a construção que contém fragmentos com baixa identidade foi identificada como pBK β -*GmSTS*-B.

2.12. Transferência dos cassetes com o promotor do gene da subunidade α da β -conglucininina do vetor pBKN para o vetor pCAMBIA 3301

Após a confirmação dos fragmentos *sense* e *antisense* no vetor pBKN, os dois cassetes (promotor do gene da subunidade α da β -conglucininina + fragmento *sense* + intron PDK + fragmento *antisense* + terminador OCS) foram liberados deste vetor e clonados no vetor binário pCAMBIA 3301.

Os procedimentos de clivagem, purificação, reação de ligação, transformação e confirmação da clonagem foram conduzidos conforme descrito para a transferência das construções com o promotor 35SCaMV, do vetor pKANNIBAL, para o vetor pCAMBIA 3301 (item 2.8.).

2.13. Nomenclatura dos clones com as construções dirigida pelo promotor do gene da subunidade α da β -conglucininina em vetores pCAMBIA 3301

Os clones contendo as construções dirigidas pelo promotor do gene da subunidade α da β -conglucininina em vetor pCAMBIA 3301 também serão denominados de acordo com o vetor usado (pCAMBIA), o nome do promotor de forma abreviada (β), e as regiões de interesse do gene *GmSTS* para o silenciamento. Dessa maneira as nomenclaturas utilizadas para as construções foram as seguintes: pCB β -GmSTS-A e pCB β -GmSTS-B, para regiões de alta e baixa identidade com os genes rafinose sintases de soja, respectivamente.

2.14. Clonagem do gene estaquiase sintase completo

O RNA total foi isolado a partir de sementes de soja no 7º estágio e primeira fita de cDNA foi sintetizada usando o *Kit SuperScript™ First-Strand Synthesis System* (INVITROGEN) em combinação com oligonucleotídeos iniciadores Oligo(dT) conforme descrito no item 2.2. Para amplificação do cDNA completo, oligonucleotídeos iniciadores gene-específicos foram construídos baseados no genoma da soja, recentemente liberado (Schmutz et al. 2010), introduzindo um sítio de restrição para *Nde I* na extremidade 5' (5'-TAATCATATGTTTCCATGGCTCCTCCAA-3') antes ao ATG e um sítio de restrição para *Xho I* (5'-TAATCTCGAGGGTAGAACCACAGGACAAAA-3') na seqüência do códon de parada na região codificadora para posteriores subclonagens. PCR foi realizado com 1 μ L de cDNA em um volume total de 25 μ L, contendo tampão IB 10X (Tris pH 8,4 100 mM; KCl 500 mM; Triton X-100 1%; MgCl₂ 15 mM), 0,2 μ M de oligonucleotídeos iniciadores, dNTP's 0,2 mM, 1U de *Taq* DNA Polimerase (PHONEUTRIA). Após a desnaturação inicial a 94°C por 3 min, as misturas de reações foram submetidas a 40 ciclos a 94°C por 1 min, 63°C por 1 min e 72°C por 2 min, seguido por uma incubação final de 10 min a 72°C. O produto de PCR foi purificado do gel de agarose (1%) com o

kit Wizard[®] *SV Gel PCR Clean-Up System* (PROMEGA), ligado ao vetor pGEM T-*Easy*, e o diagnóstico dos transformantes foi realizado por PCR e restrição enzimática. O sequenciamento foi realizado pela empresa Macrogen (Seoul, Coreia do Sul) e as sequências obtidas foram analisadas e editadas pelo programa computacional *Sequencher* versão 4.1.4, sendo o alinhamento com outras sequências disponíveis no *Genbank* realizado através do algoritmo *blastn* e do programa ClustalW (Thompson et al., 1994).

3. RESULTADOS E DISCUSSÃO

3.1. Isolamento e clonagem do gene estaquiase sintase

Com a recente liberação do genoma completo da soja (Schmutz et al., 2010), foi possível o isolamento do gene *GmSTS* completo. Um fragmento de 2.609 pb foi amplificado a partir do cDNA de sementes de soja, usando oligonucleotídeos iniciadores específicos, clonado no vetor pGEM T-*Easy* e inserido em células de *E. coli* DH5 α . Este clone foi designado de pGEM-GmSTS. Após confirmação por PCR e restrição enzimática (Figura 1) este clone foi sequenciado. A seqüência de aminoácido deduzida (Figura 2) revelou uma identidade de 99% com o gene da estaquiase sintase identificado no genoma da soja (*Glyma19g40550*), variedade Williams 82, disponível no Phytozome GBrowser (<http://www.phytozome.net/soybean>) (Anexo 1). Quando comparada com outras estaquiase sintases de plantas caracterizadas, a seqüência de aminoácidos apresentou maior identidade com estaquiase sintase de feijão adzuki (*Vigna angularis* - CAB64363) equivalente a 86%.

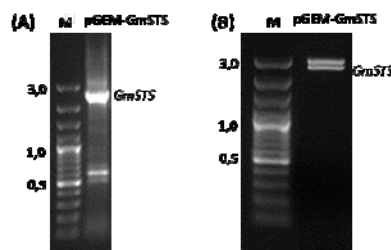


Figura 1 Ensaios para a confirmação da clonagem do cDNA do gene *GmSTS* de soja no vetor pGEM-T *Easy*. (A) PCR realizada com oligonucleotídeos iniciadores específicos; (B) Restrição enzimática com *Nde I* e *Xho I*. (M) Representa o marcador de tamanho molecular de 100 pb (FERMENTAS). Os tamanhos correspondentes estão indicados em kb. O clone foi designado de pGEM-GmSTS.

TTTCCATGGCTCCTCCAAAACAATCCAGTGAATCCACCTTGGGATTTAAGTCACTGGAA 59
M A P P N N P V N S T L G F K S L E
60 AAGGTTTTTGATTTATGTGACGGGAAATTCACCGTGAGAGGTGTGCCATTACTCTCTCAG 119
19 K V F D L C D G K F T V R G V P L L S Q
120 GTACCCAATAACGTCACTTTTCAGTTCCTTCTCCTCAATCTGTGAACCCCGTGACGCCCA 179
39 V P N N V T F S S F S S I C E P R D A P
180 CCTTCCATCCTTCAACGTGTCATTGCCGTGTCACACAAAGGAGGCTTCTTCGGATTCTCC 239
59 P S I L Q R V I A V S H K G G F F G F S
240 CAGGTTTTCTCCCTCCGACAGATTAACAAACTCCTTAGGTAGTTTCAGTGGCAGAAACTTC 299
79 Q V S P S D R L T N S L G S F S G R N F
300 CTTAGCATCTTCAGGTTCAAACATGGTGGTCTACCCAGTGGGTAGGAAATTCGGTTCA 359
99 L S I F R F K T W W S T Q W V G N S G S
360 GACCTTCAAATGGAACCCAGTGGTCTCATAGAAATCCCCGAAATAAAATCCTATGTT 419
119 D L Q M E T Q W V L I E I P E I K S Y V
420 GTAATCATTCCCATCATTGAAAAATCTTTCAGGTCCGCACTTCCCTGGCTCTGATGGC 479
139 V I I P I I E K S F R S A L H P G S D G
480 CATGTCATGATTTGTGCTGAGAGTGGTCAACTCAAGTGAAGCATCGAGTTTCGGTGCA 539
159 H V M I C A E S G S T Q V K A S S F G A
540 ATTGCTTATGTCCACGTTTCTGAGAACCCTTACAACGTGATGAAAGAAGCCTATAGTGTT 599
179 I A Y V H V S E N P Y N V M K E A Y S V
600 CTCAGGGTTCACCTCGATTGCTTTCAGGTTGTTGGAGGAGAAAACGGTCCAAAAATGCT 659
199 L R V H L D S F R L L E E K T V P K I A
660 GACAAGTTCCGTTGGTGCCTTGGGATGCGTTCTACTTAACCGTGAACCCGTTGGGGTT 719
219 D K F G W C T W D A F Y L T V N P V G V
720 TGGCATGGGCTTAAGGATTTTGCCGAGGGTGGGGTGGCTCCGAGGTTTGTATCATTGAT 779
239 W H G L K D F A E G G V A P R F V I I D
780 GATGGCTGGCAAAGTGTGAATTTTGATGGTGTGATGACCCCAACGTGGATGCTAAGAATCTT 839
259 D G W Q S V N F D G D D P N V D A K N L
840 GTTCTTGCGGGGAACAAATGACTGCGAGGCTTCATAGATTTGAAGAATGTGACAAGTTT 899
279 V L G G E Q M T A R L H R F E E C D K F
900 GGAAGTTACCAAAGGGGCTCCTTTTGGGTCTAATGCTCCTTCTTTTAACCCAAAGACG 959
299 G S Y Q K G L L L G P N A P S F N P K T
960 **GTTAAGGAGTTGATTGCGAAGGGGATAGAAGTTGAGCGTTTGGGGAAGCTGCGTGATGAG** 1019
319 **V K E L I A K G I E V E R L G K L R D E**
1020 **GCTGTTTCGTTCTGGGGTTTCTGATTGAGTTTGACCGAGATTGAGTCGAGGATTGTAAG** 1079
339 **A V S S G V S D L S L T E I E S R I V K**
1080 **GTAAAAAGGAAATTGATGATCTCTTTGGTGGGAGGGAAAGGAGAACAAGAATTATGT** 1139
359 **V K K E I D D L F G G E G K E N K E L C**
1140 **GGAGGGTGTGTGCAAAACAAATGAGTGTGGTGGGATTAAGGCTTTCATAAGGGACTTG** 1199
379 **G G C C C K T N E C G G I K A F I R D L**
1200 AGGACTGAATTCAAAGGTTTGGATGATGTCTATGTGTGGCATGCCCTTTGTGGCTCGTG 1259
399 R T E F K G L D D V Y V W H A L C G S W
1260 GGTGGTGTGAGGCCAGGACCCACACTTGAATTCAAAATAACACCTTGCAAACTCTCC 1319
419 G G V R P G A T H L N S K I T P C K L S
1320 CCTGGCCTTGATGGGACCATGCAAGATCTTGCTGTGGTTAAAATAGTGAAGGTTCCATA 1379
439 P G L D G T M Q D L A V V K I V E G S I
1380 GGACTTGTTTCATCCTGATCAAGCTAATGACTTGACGATTCCATGCCTTCTATCTTGCC 1439
459 G L V H P D Q A N D L Y D S M H S Y L A
1440 CAATCTGGTGTACCGGAGTCAAATGACGCTTTTCATAGTCTTGAATATGTGTGCGAG 1499
479 Q S G V T G V K I D V F H S L E Y V C E
1500 GAATATGGAGCAGAGTCGAGCTTGCAAAGGCTTATTACGATGGGTTGACAACTCTATT 1559
499 E Y G G R V E L A K A Y Y D G L T N S I
1560 GTCAAGAATTTAATGGAAGTGAATCATCGCTAGCATGCAGCAGTGAACGACTTTTTTC 1619

```

519  V K N F N G S G I I A S M Q Q C N D F F
1620  TTCCTTGAACCAAGCAAATTCCCCATGGGAAGAGTGGGGATGACTTTTGGTCCAAGAC 1679
539  F L G T K Q I P M G R V G D D F W F Q D
1680  CCCAATGGGGACCAATGGGAGTGTCTGGTTACAAGGGGTGCACATGATTCCGCTGTGCC 1739
559  P N G D P M G V F W L Q G V H M I R C A
1740  TACAACAGTTTGTGGATGGGGCAGATGATTCAAGCCGATTGGGACATGTTCCAATCGGAT 1799
579  Y N S L W M G Q M I Q P D W D M F Q S D
1800  CATGTGTGTGCCAAATTTTCATGCGGGTTCGAGGGCTATTTGTGGCGGTCCCTGTCTATGTA 1859
599  H V C A K F H A G S R A I C G G P V Y V
1860  AGTGACAGTGTGGGCTCTCATGACTTTGATCTCATTAAGATGCTGTGTTCCTGATGGT 1919
619  S D S V G S H D F D L I K M L V F P D G
1920  ACCGTGCCCAATGCATACATTTTGCACCTTCCAACAAGAGATTGCCTTTTCAAGAACCCT 1979
639  T V P K C I H F A L P T R D C L F K N P
1980  CTCTTTGACCAAAAACCGTTCTCAAAATTTGGAACCTCAACAAGTATGGAGGAGTTATT 2039
659  L F D Q K T V L K I W N F N K Y G G V I
2040  GGTGCTTCCAACGTCAAGGGGTGGTTGGGACCCTAAGATGAAGAAGATCAAGGGTTTC 2099
679  G A S N C Q G A G W D P K M K K I K G F
2300  TCTGAATGCTACAGGCCAATTTCTTGTACTGTGCATGTAAGTGAATGGGACCAA 2159
699  S E C Y R P I S C T V H V T E V E W D Q
2160  AAGAAAGAAGCAGTTCACATGGGTAAGGCAGAGGAGTATGTCGTGTATCTCAATCAGGCT 2219
719  K K E A V H M G K A E E Y V V Y L N Q A
2220  GAGGAACGTGCAATTCATGACCCCAAAGTCTGAACCGCTCCAATTTACTATTCAACCTTCC 2279
793  E E L H F M T P K S E P L Q F T I Q P S
2280  ACTTTTGAGATCTACAACCTTGTCCCAGTTGAAAAGCTAGGTGGCAGCATCAAATTTGCA 2339
759  T F E I Y N F V P V E K L G G S I K F A
2340  CCAATTGGGCTCACAAACATGTTCAACAGTGGAGGGACAATTCAAGAATTGGAGTGTGTT 2399
779  P I G L T N M F N S G G T I Q E L E C V
2400  GAGAAGGGTGCAAAGGTTAAGGTTAAGGGTGTGGGAGATTCCCTTGCTTACTCAAGTGAA 2459
799  E K G A K V K V K G D G R F L A Y S S E
2460  TCTCAAAGAAGTTCCAACTGAATGGTCTGATGTTGCTTTTGAGTGGCTCCCTGATGGA 2519
819  S P K K F Q L N G S D V A F E W L P D G
2520  AACTCACTCTCAACCTTGCTTGGATTGAAGAGAATGGCGGGTTTCTGATTGGCAATT 2579
839  K L T L N L A W I E E N G G V S D L A I
2580  TTCTTCTAGGTTTTCCTGTGGTTCTACC 2609
859  F F *

```

Figura 2 Seqüência nucleotídica completa e seqüência deduzida de aminoácidos do cDNA *GmSTS* clonado. A metionina no início da seqüência da proteína STS é marcado pela impressão inversa. As regiões sublinhadas são seqüências de anelamento dos oligonucleotídeos iniciadores para amplificação do gene inteiro. Os traços indicam as sequencias de anelamento dos oligonucleotídeos iniciadores específicos para amplificação dos fragmentos *sense* e *antisense*. A seqüência assinatura de estaquiase sintase, com cerca de 80 resíduos de aminoácido, está destacada em negrito.

A seqüência de nucleotídeos clonada demonstrou conter uma pequena região 5' não traduzida (UTR) e uma região 3' UTR, além de uma ORF (*Open Read Frame*) de 2.583 pb codificando um polipeptídeo de 860 aminoácidos. Com base na organização genômica da soja, o cDNA isolado parece estar presente dentro de um único cromossomo (cromossomo 19). Embora ele pertença a uma região duplicada no cromossomo 19, a presença de uma seqüência de aproximadamente 80 resíduos de

aminoácidos (Figura 2), que tem sido descrita como assinatura de proteínas estaquiase sintases (Peterbauer et al., 1999; Peterbauer et al., 2002), o distingue dos genes que codificam para rafinose sintases em soja (dados não mostrados) e sugere presença de apenas um gene que codifica a enzima STS.

Considerando a complexidade do genoma da soja, a existência de uma única cópia pode facilitar o silenciamento do gene da estaquiase sintase. Tem sido descrito que genes de cópia única são eficazmente silenciados em plantas (Miki e Shimamoto, 2004; Miki et al., 2005; Gould e Kramer, 2007; Warthmann et al. 2008). O silenciamento do gene cópia única que codifica para fitoeno desaturase (PDS), enzima que catalisa as reações iniciais na conversão do fitoeno (precursor incolor dos carotenóides) em carotenóides coloridos, tem sido bastante estudado em arroz por resultar em um fenótipo de fácil identificação que é a alteração da cor do grão. O silenciamento do gene *PDS*, por *hairpin* RNA (hpRNA), resultou em um forte silenciamento gênico e um fenótipo albino em arroz (Miki e Shimamoto, 2004; Miki et al., 2005). Em estudos posteriores com diferentes variedades de arroz (Nipponbare e IR64), o gene *PDS* também foi silenciado, através do uso de artificial miRNA (amiRNA), gerando cerca de 90% de plantas Nipponbare transgênicas e 56% de plantas IR64 transgênicas com fenótipo albino (Warthmann et al. 2008). Em aquilégia (*Aquilegia vulgaris*), uma planta herbácea, o silenciamento do gene cópia-única que codifica para antocianidina sintase (ANS), enzima que catalisa o penúltimo passo da biossíntese de antocianinas, via silenciamento gênico induzido por vírus (*Virus-Induced Gene Silencing* - VIGS) reduziu a conversão de leucoantocianidinas (incolor) em antocianida, inibindo o desenvolvimento da cor roxa da flor do tipo selvagem. Em flores com fenótipos albinos a expressão do gene *ANS* demonstrou ser significativamente reduzida (Gould e Kramer, 2007).

Embora o gene *GmSTS* clonado demonstre ter alta identidade com outros genes *STS* de plantas já descritos, para confirmação, este gene de soja está sendo expresso em *Pichia pastoris* para testes funcionais.

3.2. Construção de cassetes de expressão para o silenciamento gênico da estaquiase sintase em soja

Embora genes cópias únicas possam ser fortemente silenciados, a escolha da região a ser utilizada para gerar a repetição invertida da sequência de interesse no

cassete de expressão tem demonstrado também ser importante para garantir a eficiência do processo de silenciamento (Yamada et al., 2007). O uso de regiões conservadas entre os membros de uma família multigênica tem demonstrado ser capaz de suprimir a expressão de todos os membros da família gênica em diferentes espécies de plantas (Miki et al., 2005; Sunilkumar et al., 2006; Flores et al., 2008). Em arroz, a escolha de uma região altamente conservada, entre os genes que codificam para GTPase, a ser usada como repetição invertida em cassetes foi associada com a supressão de todos os sete genes desta família multigênica (Miki et al., 2005). Flores et al. (2008) obtiveram linhagens de soja transformadas nas quais os três membros da família *FAD3*, que codificam para enzimas que catalisam a síntese de ácido α -linolênico, foram silenciados; o alto nível de silenciamento conseguido foi atribuído ao uso de um domínio altamente conservado na região codificadora dos três genes *FAD3* como repetição invertida. Em sementes de algodão, a expressão de hpRNA, construídos a partir de uma seqüência altamente conservada entre os genes que codificam para δ -candinine sintase, que catalisa a síntese de um dos precursores do gossipol (terpenóide tóxico) resultou claramente na supressão dos transcritos de δ -candinine sintase em embriões transgênicos, além da considerável redução dos níveis de gossipol em sementes transgênicas (Sunilkumar et al., 2006).

No nosso estudo, as regiões escolhidas para produzir as repetições invertidas foram realizadas a partir de um fragmento do gene da estaquiose sintase previamente isolado (Fialho, 2007), pois o genoma da soja ainda não havia sido sequenciado completamente e o gene que codifica a enzima estaquiose sintase em soja ainda não havia sido identificado. Sendo assim, uma região do gene estaquiose sintase com baixa identidade de seqüência com os genes rafinose sintases, de 387 pb, foi escolhida para produzir repetições invertidas nos cassetes hpRNA. Também foi usada uma região conservada entre os genes estaquiose e rafinose sintases de soja, de 366 pb, para gerar repetições invertidas em construções adicionais. O uso de ambas as regiões como repetição invertida nos cassetes hpRNA pode restringir ou ampliar o efeito do silenciamento do gene que codifica para estaquiose sintase nas plantas de soja.

A clonagem dos fragmentos correspondentes as regiões de interesse do gene *GmSTS* foi realizada a partir de cDNA de sementes maduras de soja. Os tamanhos dos fragmentos amplificados corresponderam aos tamanhos esperados para os produtos de PCR a partir dos oligonucleotídeos iniciadores construídos para este trabalho. Aproximadamente 200-500 pb das seqüências alvo têm sido clonadas em direções *sense*

e *antisense*, para gerar um silenciamento eficaz (Ogita et al., 2004; Miles et al., 2005; Yamada et al., 2007; Flores et al., 2008; Wang e Xu, 2008). O uso de fragmentos menores favorece a eficiência de clonagem e transformação uma vez que a construção de cassetes de expressão envolve vetores de alto peso molecular.

Os fragmentos *sense* e *antisense* amplificados a partir de cada região de interesse do gene *GmSTS* foram purificados e clonados individualmente no vetor pGEM T-*Easy*. O diagnóstico dos transformantes foi realizado por PCR (Figura 3A) e clivagem enzimática (Figura 3B). Os clones foram confirmados por sequenciamento a partir dos DNA plasmidiais purificados.

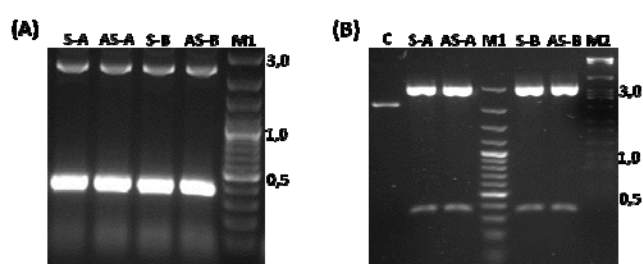


Figura 3 Ensaios para a confirmação da clonagem dos fragmentos *sense* e *antisense* das diferentes regiões do gene estaquiose sintase no vetor pGEM T-*Easy*. (A) Representa a reação de PCR; (B) Representa a reação de clivagem com DNA plasmidial dos clones isolados. Os fragmentos *sense* e *antisense* originados da região do gene estaquiose sintase com alta identidade de sequência com genes rafinose sintases estão indicados como S-A, AS-A, respectivamente. Os fragmentos *sense* e *antisense* correspondentes a região do gene estaquiose sintase com baixa identidade de sequência com genes rafinose sintases estão representados como S-B e AS-B. (M1) Representa o marcador de tamanho molecular de 100 pb (FERMENTAS); (M2) Representa o DNA de fago λ clivado com as enzimas *Eco* RI, *Bam* HI e *Hind* III, usado como marcador de tamanho molecular. Os tamanhos correspondentes estão indicados em kb. (C) Corresponde ao clone não clivado. Os clones foram designado de pGEM-GmSTSS-A, pGEM-GmSTSAS-A, pGEM-GmSTSS-B e pGEM-GmSTSAS-B.

Os fragmentos apresentaram 366 pb (S-A; AS-A) e 387 pb (S-B; AS-B). Cada uma das sequências obtidas foi alinhada com a sequência do gene estaquiose sintase (*Glyma19g40550*) disponível no Phytozome GBrowser confirmando a identidade das mesmas (Figura 4). Análises por *blastn* contra a coleção de nucleotídeos foram realizadas reforçando a existência de similaridade entre os fragmentos clonados com as sequências dos genes de estaquiose sintases de plantas já caracterizados e disponíveis no Genbank.

Para construção dos cassetes de expressão, os fragmentos *sense* foram, posteriormente, liberados do vetor pGEM T-*Easy* através de clivagem enzimática usando as enzimas de restrição *Xho* I e *Kpn* I, purificados e ligados aos vetores pKANNIBAL e pBKN. O diagnóstico dos transformantes foi realizado por PCR e clivagem enzimática a partir do DNA plasmidial dos clones isolados. Os clones positivos que apresentavam fragmento *sense* pertencentes à região de alta identidade com as sequências de rafinose sintases de soja foram sequenciados e designados de pK35S-GmSTS-SA e pBK β -GmSTS-SA, quando ligados aos vetores pKANNIBAL e pBKN, respectivamente (Figura 5). Semelhantemente, os clones positivos contendo o fragmento *sense* pertencentes à região de baixa identidades de sequências com genes rafinose sintases foram designados de pK35S-GmSTS-SB, quando clonado nos vetores pKANNIBAL, e pBK β -GmSTS-SB, quando ligado ao vetor pBKN (Figura 5).

Para completar a construção dos cassetes, o fragmento *antisense* cuja seqüência corresponde a região de alta identidade com genes rafinose sintase foram liberados do vetor pGEM T-*Easy*, por clivagem enzimática com as enzimas *Cla* I e *Xba* I, e inseridos nos clones de pK35S-GmSTS-SA e pBK β -GmSTS-AS anteriormente obtidos. Os clones positivos foram identificados por PCR e clivagem enzimática como demonstrado nas Figuras 5 e 6. Após sequenciamento para confirmação da identidade das sequências dos fragmentos *antisense*, os clones obtidos, contendo ambos os fragmentos, foram designados como pK35S-GmSTS-A e pBK β -GmSTS-A. O fragmento *antisense* com baixa identidade de seqüência com genes rafinose sintases foram inseridos nos clones pK35S-GmSTS-SB e pBK β -GmSTS-SB, após serem liberados do vetor pGEM T-*Easy* com as mesmas enzimas de restrição. Os clones confirmados por PCR, sequenciamento e/ou clivagem enzimática (Figuras 6) foram denominados pK35S-GmSTS-B e pBK β -GmSTS-B. Os fragmentos *antisense* dos clones contendo o promotor da subunidade α da proteína β -conglucina, pBK β -GmSTS-A e pBK β -GmSTS-B, embora confirmados por PCR (Figura 5), não puderam ser liberados por clivagem enzimática. A troca do promotor original (35SCaMV) do vetor pKANNIBAL pelo promotor semente-específico, juntamente com a inserção de sequências repetidas e invertidas podem ter resultado na formação de estruturas cruciformes entre as sequências clonadas, impedindo o acesso das enzimas de restrição aos seus sítios correspondentes. Sendo assim, neste caso, a confirmação da clonagem foi realizada por sequenciamento a partir dos fragmentos *antisense* amplificados e purificados.

(A)

Glyma19g40550	GGAGGCAGAGTCGAGCTTGCAAAGGCTTATTACGATGGGTTGACAAACTCTATTGTCAAG	1560
Clone A	-----CGAGCTTGCAAAGGCTTATTACGATGGGTTGACAAACTCTATTGTCAAG	49
Glyma19g40550	AATTTAATGGAAGTGAATCATCGCTAGCATGCAGCAGTGCAACGACTTTTCTTCCTT	1620
Clone A	AATTTAATGGAAGTGAATCATCGCTAGCATGCAGCAGTGCAACGACTTTTCTTCCTT	109
Glyma19g40550	GGAACCAAGCAAATCCCATGGGAAGAGTTGGGGATGACTTTTGGTTCCAAGACCCCAAT	1680
Clone A	GGAACCAAGCAAATCCCATGGGAAGATTTGGGGATGACTTTTGGTTCCAAGACCCCAAT	169
Glyma19g40550	GGGACCCAATGGGAGTGTCTGGTTACAAGGGCTGCACATGATTCACATGTCCTACAAC	1740
Clone A	GGGACCCAATGGGAGTGTCTGGTTACAAGGGCTGCACATGATTCACATGTCCTACAAC	228
Glyma19g40550	AGTTTGTGGATGGGCAGATGATTCAGCCCATTGGGACATGTTCCAATCGGATCATGTG	1800
Clone A	AGTTTGTGGATGGGCAGATGATTCAGCCCATTGGGACATGTTCCAATCGGATCATGTG	288
Glyma19g40550	TGTCGCCAAATTCATGCGGGTTTCGAGGGCTATTTGTGGCGTCTGTCTATGTAAGTGA	1859
Clone A	TGTCGCCAAATTCATGCGAGTTTCGAGGGCTATTTGTGGCGTCTGTCTATGTAAGTGA	348
Glyma19g40550	CAGTGTGGGCTCTCATGACTTTGATCTCATTAAAGATGCTTGTGTTCCCTGATGGTACCGT	1919
Clone A	CAGTGTGGGCTCTCATGA-----	366

(B)

Glyma19g40550	TCAGGGTTCACCTCGATTTCGTTTCAGGTTGTTGGAGGAGAAAAACCGTGCCAAAAATTGCTG	660
Clone B	-----TGCCAAAAATTGCTG	15
Glyma19g40550	ACAAGTTCGGTTGGTGCACCTGGGATGCGCTTCTACTTAACCGTGAACCCGTGTTGGGGTTT	720
Clone B	ACAAGTTCGGTTGGTGCACCTGGGATGCGCTTCTACTTAACCGTGAACCCGTGTTGGGGTTT	75
Glyma19g40550	GGCATGGGCTTAAGGATTTTGCCGAGGGTGGGGTGGCTCCGAGGTTTGTATCATTGATG	780
Clone B	GGCATGGGCTTAAGGATTTTGCCGAGGGTGGGGTGGCTCCGAGGTTTGTATCATTGATG	135
Glyma19g40550	ATGGCTGGCAAAGTGTGAATTTTGATGGTGTGACCCCAACGTGGATGCTAAGAATCTTG	840
Clone B	ATGGCTGGCAAAGTGTGAATTTTGATGGTGTGACCCCAACGTGGATGCTAAGAATCTTG	195
Glyma19g40550	TTCTTGCGGGGAACAAATGACTGCGAGGCTTCATAGATTTGAAGAATGTGACAAGTTTG	900
Clone B	TTCTTGCGGGGAACAAATGACTGCGAGGCTTCATAGATTTGAAGAATGTGACAAGTTTG	255
Glyma19g40550	GAAGTTACCAAAGGGGCTCCTTTGGGTCCTAATGCTCCTCTTTTAACCCAAAGACGG	960
Clone B	GAAGTTACCAAAGGGGCTCCTTTGGGTCCTAATGCTCCTCTTTTAACCCAAAGACGG	315
Glyma19g40550	TTAAGGAGTTGATTGCGAAGGGGATAGAAGTTGAGCGTTTGGGGAAGCTGCGTGATGAGG	1020
Clone B	TTAAGGAGTTGATTGCGAAGGGGATAGAAGTTGAGCGTTTGGGGAAGCTGCGTGATGAGG	375
Glyma19g40550	CTGTTTCGCTCGGGTTTCTGATTTGAGTTGACCGAGATTGAGTCGAGGATTGTGAAGG	1080
Clone B	CTGTTTCGCTCG-----	387

Figura 4 Alinhamento entre as sequências dos fragmentos clonados e a sequência do gene de estaquiose sintase (*Glyma19g40550*) disponível no Phytozome GBrowser. Os Clones A e B correspondem a sequências dos fragmentos com alta e baixa identidade com as sequências dos genes da rafinose sintases em soja. O Clone B corresponde a sequências dos fragmentos com baixa identidade com as sequências dos genes da rafinose sintases em soja, respectivamente. Bases alteradas estão marcadas em cinza e deleções em azul.

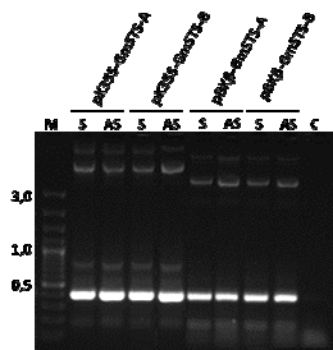


Figura 5 Reação de PCR para confirmação da clonagem dos fragmentos *sense* e *antisense* correspondente a cada região de interesse do gene estaquiase sintase nos vetores pKANNIBAL (pK35S-GmSTS-A e pK35S-GmSTS-B) e pBKN (pBKβ-GmSTS-A e pBKβ-GmSTS-B). Os fragmentos *sense* e *antisense* estão representados como S e AS, respectivamente. O controle negativo da reação (sem DNA) está representado por (C); (M) corresponde ao marcador de tamanho molecular de 100 pb (FERMENTAS).

Os clones pK35S-GmSTS-A, pK35S-GmSTS-B, pBKβ-GmSTS-A, pBKβ-GmSTS-B contêm os cassetes com sequências repetidas e invertidas do gene estaquiase sintase que proporcionam a formação do hpRNA, resultando na estrutura necessária para o silenciamento deste gene em soja. Entretanto os cassetes construídos não possuem uma marca para seleção dos transgenes em plantas. Além disso, como a transformação de plantas será realizada, a princípio, por meio de *Agrobacterium tumefaciens*, uma estrutura de T-DNA também se faz necessária. Assim, todas as construções obtidas foram liberadas dos clones pK35S-GmSTS-A, pK35S-GmSTS-B, pBKβ-GmSTS-A e pBKβ-GmSTS-B, com o uso das enzimas de restrição *Pst* I e *Sac* I, e inseridas no vetor binário pCAMBIA 3301. A clonagem foi confirmada por PCR e clivagem enzimática e os clones confirmados contendo fragmentos do gene estaquiase sintase com alta identidade de seqüência com os genes rafinose sintases foram denominados de pCB35S-GmSTS-A e pCBβ-GmSTS-A, enquanto os clones apresentando fragmentos com baixa identidade de seqüência com os genes rafinose sintase foram designados pCB35S-GmSTS-Be pCBβ-GmSTS-B (Figura 7).

As construções hpRNA obtidas e inseridas no vetor pCAMBIA 3301, para serem posteriormente usadas para transformação genética de soja via *Agrobacterium tumefaciens*, estão representadas de maneira esquemática nas figura 8 e 9.

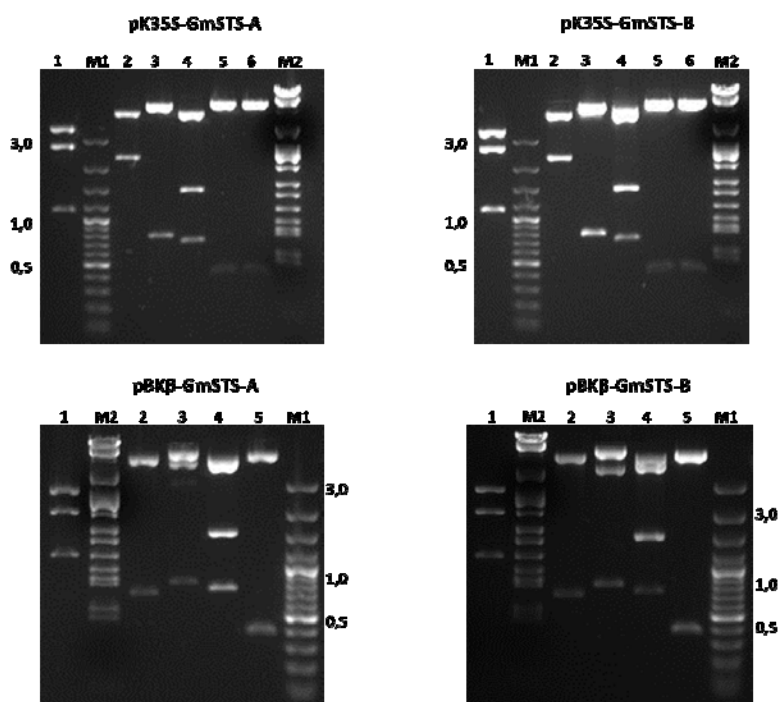


Figura 6 Reação de clivagem para confirmação das clonagens dos fragmentos *sense* e *antisense* correspondentes as regiões de interesse no gene estaquiose sintase no vetor pKANNIBAL e pBK β . As canaletas correspondem à liberação da construção inteira (1), do promotor (2), do intron (3), do terminador (4), do fragmento *sense* (5) e do fragmento *antisense* (6). (M1) Representa o marcador de tamanho de 100 pb (INVITROGEN); (M2) Representa o DNA do fago λ clivado com as enzimas *Eco* RI, *Bam* HI e *Hind* III, utilizado como marcador de tamanho molecular.

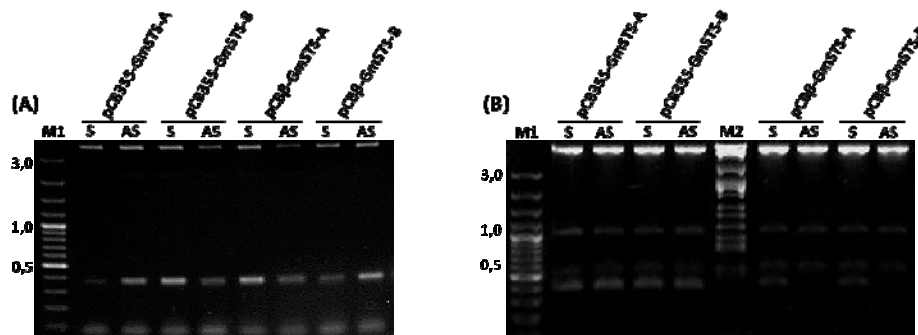


Figura 7 Ensaio para confirmação da clonagem dos fragmentos *sense* e *antisense* das diferentes regiões de interesse no gene estaquiose sintase no vetor pCAMBIA 3301. (A) Representa a reação de PCR; (B) Representa a reação de clivagem do DNA plasmidial dos clones isolados. Os fragmentos *sense* e *antisense* originados da região do gene estaquiose sintase com alta identidade de sequência com genes rafinose sintases estão indicados como S-A, AS-A, respectivamente. Os fragmentos *sense* e *antisense* correspondentes a região do gene estaquiose sintase com baixa identidade de sequência com genes rafinose sintases estão representados como S-B e AS-B, respectivamente. (M1) Representa o marcador de tamanho molecular de 100 pb (FERMENTAS); (M2) Representa o DNA de fago λ clivado com as enzimas *Eco* RI, *Bam* HI e *Hind* III, usado como marcador de tamanho molecular. Os tamanhos correspondentes estão indicados em kb.

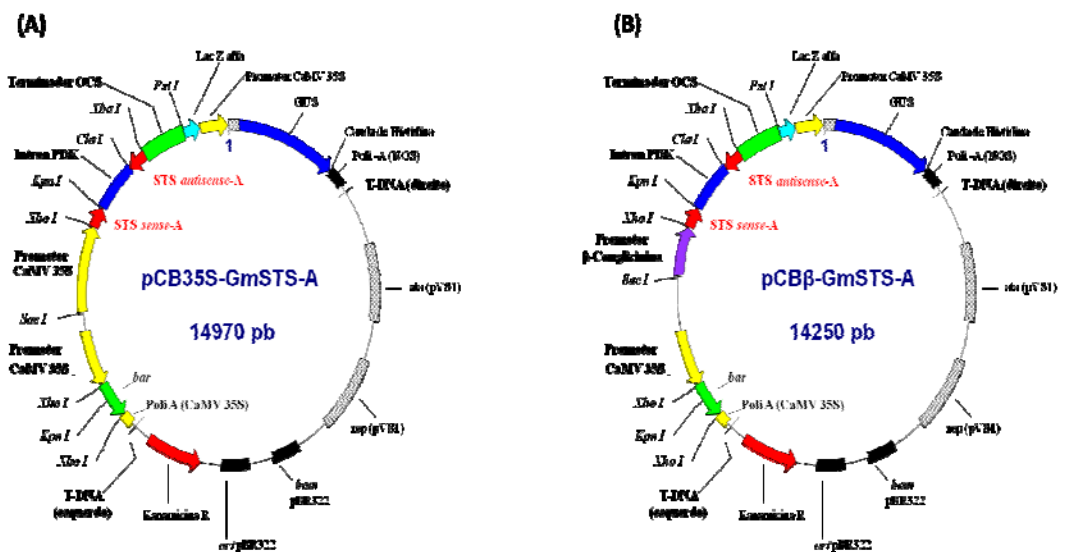


Figura 8 Representação esquemática dos clones contendo as construções para silenciamento gênico, RNA *interference*, do gene da estaquiose sintase em soja. Os fragmentos *sense* e *antisense* representado em direções opostas e em vermelhos correspondem a região do gene estaquiose sintase com alta identidade de seqüência com os genes da raffinose sintases em soja. Os cassetes estão sob controle do promotor 35SCaMV (A) e do promotor do gene da subunidade α da β -conglucina (B).

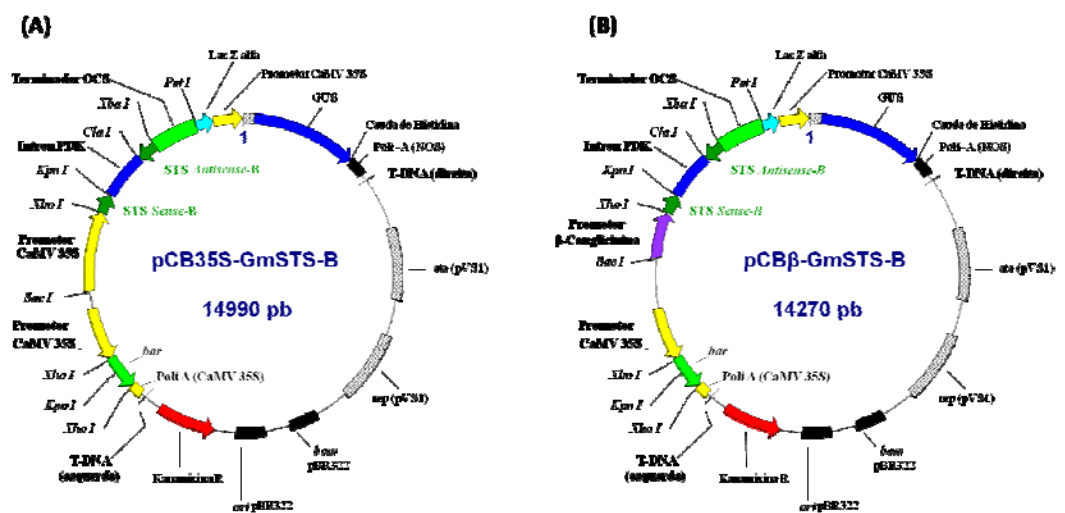


Figura 9 Representação esquemática dos clones contendo as construções para silenciamento gênico, RNA *interference*, do gene da estaquiose sintase em soja. Os fragmentos *sense* e *antisense* representado em direções opostas e em verde correspondem a região do gene estaquiose sintase com baixa identidade de seqüência com os genes da raffinose sintases em soja. Os cassetes estão sob controle do promotor 35SCaMV (A) e do promotor do gene da subunidade α da β -conglucina (B).

Construções que resultam em estruturas hpRNA têm demonstrado ser uma ferramenta consistente para reduzir significativamente um produto gênico em planta (Akashi et al., 2001; Kusaba et al., 2003; Sunilkumar et al., 2006; Nunes et al., 2006; Wang e Xu, 2008, Flores et al., 2008; Ibraheem et al., 2010). Pesquisas têm demonstrado que quase 85% (Wesley et al., 2001), 95% (Schattat et al., 2004) ou 100% (Smith et al., 2000) de plantas transformadas com construções hpRNA apresentaram genes de interesse silenciados. Nestas construções, sequências de DNA que codificam regiões auto-complementares formam uma repetição invertida (Waterhouse et al., 1998), que, após transcrição, geram dsRNA que pode ser processado pelas proteínas *DICER-like* formando pequenos RNAs que facilitam o silenciamento gênico pós-transcricional. Akashi et al., (2001) relataram que o uso de construções *-sense* e *-antisense* para silenciar o gene *FLUC*, que codifica para luciferase de vaga-lume, resultaram em uma redução da expressão gênica de apenas 25-40%; enquanto que o a utilização de cassetes hpRNA contendo repetições invertidas proporcionaram uma redução de 85% da expressão do mesmo gene, usando a mesma seqüência, em suspensão de células de tabaco. Cartea et al. (1998) estudaram o efeito do silenciamento gênico do gene *FAD2-1*, que codifica a enzima ω -6 dessaturase microssomal em *Arabidopsis*, usando construções *-sense* e *-antisense*. A construção *sense* promoveu uma taxa de silenciamento de 10% entre as 41 plantas transgênicas obtidas, enquanto a construção *antisense* produziu aproximadamente 15% das plantas transgênicas exibindo silenciamento. Estes resultados são bem inferiores se comparados aos obtidos por Singh et al. (2000), usando construções hpRNA com repetições invertidas, alcançaram aproximadamente 70% de plantas transgênica mostrando silenciamento. Neste estudo a troca do espaçador entre as repetições invertidas por um intron também resultou em um aumento da eficiência de silenciamento, sendo que, neste caso, 100% das plantas transgênicas regeneradas demonstraram o silenciamento do gene *FAD2-1*. Wesley et al. (2001) relataram a eficiência relacionada ao uso de construções hpRNA, co-supressão e sequências *-antisense* no silenciamento de alguns genes em plantas, incluindo o gene *CHD*, que codifica para chalcona sintase, em *Arabidopsis*. Neste estudo, a construção *sense*, usada para co-supressão, resultou em uma freqüência de silenciamento muito baixa, cerca de 10% das 19 plantas transformadas continham o gene *CHD* silenciado. A construção *antisense* não gerou plantas transgênicas silenciadas. Entretanto o uso de construção hpRNA, contendo repetições invertida do mesmo gene, espaçadas pelo intron do gene que codifica a piruvato desidrogenase quinase (PDK), no vetor

pHANNIBAL, proporcionou a obtenção de aproximadamente 90% das plantas transformadas com o gene *CHD* silenciado. A alta eficiência de silenciamento alcançada nos estudos com hpRNA indicam que muitas das características desejáveis no melhoramento gênico podem ser obtidas através de RNAi, especificamente pelo uso de repetições gênicas invertidas. No nosso estudo, o vetor pKANNIBAL, usado para construção do cassete de expressão para silenciamento, é semelhante ao vetor pHANNIBAL anteriormente citado, sendo que a única alteração se refere ao gene de resistência.

As variações do silenciamento gênico gerado pelo uso do mesmo tipo de construção, que vão de muito fortes a muito fracos, têm sugerido que vários fatores podem estar envolvidos ou podem influenciar o processo de RNAi. Parte do amplo espectro de variação dos efeitos do RNAi em construções hpRNA parece estar relacionada com o comprimento da seqüência de ligação entre os fragmentos *sense* e *antisense*. Curtas seqüências de ligações têm mostrado produzir um silenciamento de genes específicos mais eficientes que seqüências de ligações longas. Cassetes contendo repetições gênicas invertidas não espaçadas por uma seqüência de ligação apresentam uma eficiência de silenciamento reduzida se comparada a cassetes nos quais tais seqüências estão presentes (Singh et al., 2000; Wesley et al., 2001). A enzima 1-aminociclopropano-1-carboxilato oxidase (ACC oxidase) catalisa a oxidação de ACC a etileno; plantas transgênicas de tomate nas quais construções com repetições invertidas de uma seqüência do gene *ACC oxidase* ligadas por um longo fragmento, de 1.002 nucleotídeos, apresentaram um reduzido efeito RNAi, sendo que apenas 10% das plantas transformadas tiveram um silenciamento gênico completo (Xiong et al., 2005), embora a expressão de dsRNA tenha sido confirmada em 90% das plantas transformadas. Entretanto, quando as regiões invertidas do gene foram ligadas por um pequeno fragmento, de 7 nucleotídeos, foi observado o resultado oposto, isto é, a maioria das plantas de tomate transgênicas (70%) exibiram um silenciamento eficaz do gene *ACC oxidase* e em 30% destas plantas não houve silenciamento ou houve um reduzido silenciamento do gene em questão.

Tem sido sugerido que um fator que também pode influenciar a eficiência do silenciamento gênico é a concentração de mRNA alvo. Miki et al. (2005) relataram que o uso de uma mesma seqüência conservada promoveu o silenciamento gênico de todos os genes de uma família multigênica, que codificam uma GTPase em arroz, com eficiências variável. Esta variação foi associada com a concentração do mRNA alvo

uma vez que, dos sete genes da família, três deles (*OsRac2*, 3 e 4) são altamente expressos e eficazmente silenciados. O gene *OsRac7*, embora apresente alta homologia de seqüência com a região conservada usada como repetição invertida, foi o menos expresso e também mais fracamente silenciado. Os autores sugerem que a alta expressão dos genes *OsRac2*, *OsRac3* e *OsRac4* e o conseqüente acúmulo de transcritos pode facilitar o acesso dos siRNA e RISC a estes mRNA alvos.

A maioria dos experimentos RNAi tem usado promotores constitutivos, tais como o promotor 35S do vírus do mosaico da couve-flor (CAMV) e o promotor de gene da ubiquitina-1 (*Ubi1*) de milho, para dirigir a expressão de transgenes RNAi (Li et al., 2011). Sem restrições específicas, promotores constitutivos dirigem a alta expressão gênica em todos os tecidos/órgãos das plantas transgênicas, independentemente de reguladores tecido ou órgão específicos. Entretanto o silenciamento de genes através do uso de promotores regulados, tal como promotores tecido/órgão específicos, é interessante para alterar a composição de nutrientes em sementes. Tem sido demonstrado que os promotores específicos podem ser usados com sucesso para limitar o silenciamento em diferentes tecidos/órgãos vegetais de forma eficaz (Davuluri et al., 2005; Sunilkumar et al., 2006; Houmard et al., 2007). Transgênicos dsRNA dirigidos por promotores regulados foram usados para silenciar a expressão gênica e melhorar a qualidade nutricional de sementes de arroz (Kusaba et al., 2003), de milho (Huang et al., 2006; Houmard et al., 2007) e de frutos em tomates (Davuluri et al., 2005). Em sementes de algodão também foi demonstrado que o uso acoplado da tecnologia de RNAi com promotor tecido-específico pode reduzir significativamente e seletivamente os níveis de um composto de interesse em sementes. Sunilkumar et al., (2006) visando reduzir os níveis de gossipol em sementes de algodão, para reutilizá-las após extração das fibras na alimentação, usaram um promotor semente específico, originado do promotor do gene da subunidade α da β -globulina, para dirigir a expressão de uma construção hpRNA. Esta construção, inserida no vetor pHANNIBAL, continha seqüências repetidas e invertidas do gene que codifica σ -cadieno sintase, o qual sintetiza um precursor do gossipol. Este é um terpenóide tóxico relacionado à defesa da planta contra ataque de insetos, entretanto sua presença na semente impede a utilização da mesma como fonte de proteínas e calorias. As plantas de algodão transformadas mostraram claramente a supressão dos transcritos do gene σ -cadieno sintase e conseqüente redução dos níveis de gossipol em suas sementes; entretanto não houve

alteração do acúmulo deste terpenóide nos demais órgãos vegetais, sendo o conteúdo deste terpenóide nos outros tecidos equivalentes aos das plantas controles.

Em nosso estudo, construímos cassetes de expressão contendo duas diferentes sequências do gene estaquiose sintase que foram inseridas como repetições invertidas no vetores pKANNIBAL e pBKN. Ambos os vetores possuem um intron do gene que codifica a piruvato desidrogenase quinase (PDK) ligando os fragmentos *sense* e *antisense*, entretanto diferem quanto ao promotor que dirigirá a expressão das construções hpRNA nas plantas de soja transformadas. O vetor pKANNIBAL possui o promotor constitutivo 35SCaMV, enquanto o pBKN contém o promotor da subunidade α da proteína β -conglucininina de soja.

Ensaio de prévio de RT-PCR para análise de expressão gênica (Fialho 2007) e a análise do padrão de expressão *in silico* do gene estaquiose sintase indicam que este gene é expresso tanto em sementes, durante todos os períodos de enchimento do grão, quanto em outros órgãos vegetais (raiz, caule e folha), principalmente sobre condições de estresses biótico e abiótico. Neste caso, o uso de um promotor semente específico pode restringir o silenciamento do gene de interesse apenas à semente, uma vez que a expressão gênica e o acúmulo de estaquiose parecem estar associados a mecanismos de respostas a estresses. Entretanto, se a soja, assim como algumas espécies vegetais, utilizar rafinose e estaquiose como carboidrato de transporte, o que ainda não tem sido bem descrito, o uso de um promotor constitutivo para se reduzir ou eliminar o conteúdo de estaquiose sintase nos grãos se faz necessário. Nesse sentido, mais estudos são requeridos para se afirmar se há translocação e qual a taxa desta entre os diferentes órgãos em soja.

O número de cópias e o sítio de inserção de um transgene dentro do genoma são outros importantes fatores que parecem afetar o grau de silenciamento gênico, uma vez que influenciam o nível de expressão do mesmo. Plantas transformadas com múltiplas cópias de um transgene têm apresentado maiores tendências a serem silenciadas a nível pós-transcricional, se comparadas a plantas com um único transgene inserido (Sallaud et al., 2003; Nocarova et al., 2010). Em seus estudos Tang et al. (2007) demonstraram que a transformação de pinus com uma construção dirigida pelo promotor 35SCaMV, visando silenciar o gene *gfp* (*green fluorescent protein*) por RNAi, resultou na obtenção de 95 linhagens transgênicas com gene *gfp* silenciado contendo de 1 a 7 cópias do transgene no genoma. Ensaio de microscopia a laser e análises de *northern blot* mostraram que o silenciamento do gene de interesse foi obtido nas linhagens

transgênicas com mais de três cópias do transgene inseridas no genoma, mas não em linhagens com uma única cópia do transgene. Além disso, só foi detectado, por ensaio de proteção por RNase, siRNA de 21 nucleotídeos em linhagens transgênicas de pinus contendo número de cópias do transgene maior que três inseridas no genoma, indicando que este silenciamento do gene *gfp* pode ser associado com silenciamento gênico pós-transcricional. Entretanto, de acordo com Kerschen et al. (2004) eficiência do processo de silenciamento também depende do gene alvo considerado. Em seu estudo, o silenciamento de 25 genes endógenos diferentes em *Arabidopsis*, analisado através de RT-PCR quantitativo, revelou que linhagens independentes, homozigotas, que apresentaram a inserção de uma única cópia no genoma e que tinham o mesmo gene como alvo, apresentaram níveis de redução do transcrito na mesma extensão. Porém, linhagens que apresentaram múltiplas inserções diferiram quanto ao grau de redução da expressão do mRNA alvo, mas nunca excederam os níveis apresentados por linhagens com inserção única, sendo que a redução máxima dos transcritos variou de acordo com o gene alvo. Estas observações indicam que cada seqüência de mRNA alvo parece possuir um grau intrínseco de susceptibilidade à degradação mediada por dsRNA.

O local de inserção do transgene no genoma também afeta o silenciamento gênico. Em organismos eucariotos superiores a integração de um transgene no genoma pode ocorrer em qualquer local do genoma, incluindo heterocromatinas transcricionalmente inativas ou regiões adjacentes à heterocromatinas, quais podem bloquear ou reduzir a expressão do transgene, e conseqüentemente o silenciamento gênico pós-transcricional por hpRNA (Maessen, 1997; Butaye et al., 2005). Este mecanismo é chamado de efeito de posição e também têm sido descrito em plantas (Wisman et al., 1993; Fujiwara e Beachy, 1993).

4. CONCLUSÕES

O gene da estaquiose sintase de soja variedade comercial CAC-1 foi isolado e sequenciado. Com o genoma da soja disponível foi possível confirmar a identidade de seqüência.

Foi possível identificar e isolar duas regiões diferentes de gene da estaquiose sintase correspondente a regiões de alta (366 pb) e baixa (387 pb) homologia com os genes de rafinose sintases em soja.

Foi possível clonar os fragmentos *sense* e *antisense* pertencentes às duas distintas regiões identificadas do gene estaquiose sintase, para produzir seqüências repetidas e invertidas em construções hpRNA, em vetores pKANNIBAL e pBKN.

No total, foram obtidos quatro cassetes de expressão para silenciamento do gene, via RNA de interferência, em soja. Dois destes continham o fragmento *sense* e *antisense* cuja seqüência apresentaram alta identidade de seqüência com genes rafinose sintases de soja, enquanto os outros dois continham o fragmento *sense* e *antisense* do mesmo gene, porém com a baixa identidade de seqüência com os genes das rafinose sintases de sojas. Dois destes cassetes terão a expressão dirigida por um promotor constitutivo (35SCaMV), enquanto os dois serão dirigidos pelo promotor semente específico da subunidade α da proteína β -conglucina de soja.

5. PERSPECTIVAS

O gene da estaquiose sintase isolado será clonado e introduzido em células de *Pichia pastoris* para expressão heteróloga e posteriores ensaios funcionais.

Os cassetes obtidos serão utilizados para transformação genética de nós cotiledonares e/ou embriões zigóticos de soja, via *Agrobacterium tumefaciens*, linhagem KYTR1.

Os cassetes produzidos serão clonados no vetor pAC 321 para transformação genética de soja via biobalística, no laboratório do Dr. Francisco Aragão (EMBRAPA/CENARGEN).

As variedades de soja, a serem obtidas, serão analisadas quanto ao nível de expressão do gene da enzima estaquiose sintase e ao conteúdo de ROs em sementes de soja transformadas durante o enchimento do grão.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Akashi H, Miyagishi M, Taira K (2001) Suppression of gene expression by RNA interference in cultured plant cells. *Antisense Nucleic Acid Drug Development* 11: 359-367.
- Barros BA 2006 Construção de cassetes de expressão para o silenciamento gênico de fatores antinutricionais de soja, via interferência por RNA. Universidade Federal de Viçosa – Viçosa, MG, p.69 (Dissertação M.S.).
- Butaye KMJ, Cammue BPA, Delauré SL, De Bolle MFC (2005) Approaches to Minimize Variation of Transgene Expression in Plants. *Molecular Breeding* 16: 79-91.
- Cartea ME, Migdal M, Galle AM, Pelletier G, Guerche P (1998) Comparison of sense and antisense methodologies for modifying the fatty acid composition of *Arabidopsis thaliana* oilseed. *Plant Science* 136: 181-194.
- Davuluri GR, Van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J, Bramley PM, Pennings HM, Bowler C (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nature Biotechnology* 23: 890–895.
- Dierking EC, Bilyeu KD (2008) Association of a Soybean Raffinose Synthase Gene with Low Raffinose and Stachyose Seed Phenotype. *The Plant Genome* 2: 135-145.
- Dierking EC, Bilyeu KD (2009) Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *Journal of Plant Physiology* 166: 1329-1335.
- Fialho L da S (2007) Purificação e caracterização de uma α -galactosidase em sementes de *Tachigali multijuga* e clonagem parcial do gene da estaquiose sintase de soja. Universidade Federal de Viçosa – Viçosa, MG, p. 140. (Tese D.S.).

- Flores T, Karpova O, Su X, Zeng P, Bilyeu K, Sleper DA, Nguyen HT, Zhang ZJ (2008) Silencing of *GmFAD3* gene by siRNA leads to low α -linolenic acids (18:3) of *fad3*-mutant phenotype in soybean [*Glycine max* (Merr.)]. *Transgenic Research* 17: 839-850.
- Fujiwara T, Beachy RN (1993) Expression of Soybean Seed Storage Protein Genes in Transgenic Plants; their Effects on Expression of a Neighboring Gene and Position Dependency. *Plant Cell Physiology* 34:13-20.
- Gould B, Kramer EM (2007) Virus-induced gene silencing as a tool for functional analyses in the emerging model plant *Aquilegia* (columbine, Ranunculaceae). *Plant Methods* 3: 6.
- Houmard NM, Mainville JL, Bonin CP, Huang S, Luethy MH, Malvar TM (2007) High-lysine corn generated by endosperm-specific suppression of lysine catabolism using RNAi. *Plant Biotechnology Journal* 5: 605-14.
- Huang S, Frizzi A, Florida CA, Kruger DE, Luethy MH (2006) High lysine and high tryptophan transgenic maize resulting from the reduction of both 19- and 22-kD α -zeins. *Plant Molecular Biology* 61: 525–535.
- Ibraheem HMM, Alkharouf NW, Meyer SLF, Aly MAM, El-Din AEFYG, Hussein EHA, Matthews BF (2010) Post-transcriptional gene silencing of root-knot nematode in transformed soybean roots. *Experimental Parasitology* 127: 90-99.
- Kerschen A, Napoli CA, Jorgensen RA, Muller AE (2004) Effectiveness of RNA interference in transgenic plants. *FEBS Letters* 566: 223–228.
- Kumar V, Rani A, Goyal L, Dixit AK, Manjaya JG, Dev J, Swamy M (2010) Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *Journal of agricultural and food chemistry* 58: 5081-5085.
- Kusaba M, Miyahara K, Iida S, Fukuoka H, Takano T, Sassa H, Nishimura M, Nishio T (2003) Low glutelin content1: a dominant mutation that suppresses the glutelin multigene family via RNA silencing in rice. *Plant Cell* 15: 1455–1467.
- Leske KL, Jevne CJ, Coon CN (1993) Effect of oligosaccharide additions on nitrogen-corrected true metabolizable energy of soy protein concentrated. *Poultry Science*, 72: 664–668.
- Li J, Jiang D, Zhou H, Li F, Yang J, Hong L, Fu X, Li Z, Liu Z, Li J, Zhuang C (2011) Expression of RNA-Interference/Antisense Transgenes by the Cognate Promoters

- of Target Genes Is a Better Gene-Silencing Strategy to Study Gene Functions in Rice. *PLoS ONE* 6: e17444.
- Liu Q, Singh SP, Green AG (2002) High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiology* 129: 1732-1743.
- Maessen GDF (1997) Genomic stability and stability of expression in genetically modified plants. *Acta botanica neerlandica* 46: 3-24.
- Miki D, Itoh R, Shimamoto K (2005) RNA Silencing of Single and Multiple Members in a Gene Family of Rice. *Plant Physiology* 138: 1903-1913.
- Miki D, Shimamoto K (2004) Simple RNAi vectors for stable and transient suppression of gene function in rice. *Plant Cell Physiology* 45: 490–495.
- Miles GP, Samuel MA, Zhang Y, Ellis BE (2005) RNA interference-based (RNAi) suppression of AtMPK6, an *Arabidopsis* mitogen-activated protein kinase, results in hypersensitivity to ozone and misregulation of AtMPK3. *Environmental Pollution* 138: 230-237.
- Nocarova E, Opatry Z, Fischer L (2010) Successive silencing of tandem reporter genes in potato (*Solanum tuberosum*) over 5 years of vegetative propagation. *Annals of Botany* 106: 565-572.
- Nunes ACS, Vianna GR, Cuneo F, Amaya-Farfán J, Capdeville G, Rech EL, Aragão FJL (2006) RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (*GmMIPSI*) in transgenic soybean inhibited seed development and reduced phytate content. *Planta* 224: 125-132.
- Ogita S, Uefuji H, Morimoto M, Sano H (2004) Application of RNAi to confirm theobromine as the major intermediate for efficient biosyntheses in coffee plants with potential for construction of decaffeinated varieties. *Plant Molecular Biology* 54: 931-941.
- Parsons CM, Zhang Y, Araba M (2000) Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poultry Science* 79: 1127–1131.
- Peterbauer T, Mucha J, Mach L, Richter A (2002) Chain-elongation of raffinose in pea seeds. Isolation, characterization, and molecular cloning of multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. *Journal of Biological Chemistry* 277: 194–200.

- Peterbauer T, Mucha J, Mayer U, Popp M, Glossl J, Richter A (1999) Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*): molecular cloning and functional expression of stachyose synthase. *The Plant Journal* 20: 509-518.
- Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. *In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386
- Sallaud C, Meynard D, van Boxtel J, Gay C, Bès M, Brizard JP, Larmande P, Ortega D, Raynal M, Portefaix M, Ouwekerk PBF, Rueb S, Delseny M, Guiderdoni E (2003) Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theoretical and Applied Genetics* 106: 1396–1408.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning, a Laboratory Manual*. 2nd ed. USA: Cold Spring Harbor Laboratory.
- Schattat MH, Klösgen RB, Marques JP (2004) A Novel Vector for Efficient Gene Silencing in Plants. *Plant Molecular Biology Reporter* 22: 145–153.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean. *Nature* 463: 178-183.
- Singh S, Green A, Stoutjesdijk P, Liu Q (2000) Inverted-repeat DNA: a new gene-silencing tool for seed lipid modification. *Biochemical Society Transactions* 28: 925-927.
- Smiricky HR, Grieshop CH, Albin DM, Wubben JE, Gabert VM, Fahey GCJ (2002) The influence of soy oligosaccharides on apparent and true ideal amino acid digestibilities and fecal consistency in growing pigs. *Journal of Animal Science* 80: 2433-2441.
- Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG, Waterhouse PM (2000) Total silencing by intron-spliced hairpin RNAs. *Nature* 407: 319–320.
- Suarez FL, Springfield J, Furne JK, Lohrmann TT, Kerr PS, Levit MD (1999) Gas production in humans ingesting a soybean flour derived from beans naturally low in oligosaccharides. *American Journal of Clinical Nutrition* 69:135–139.

- Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic. *Proceedings of the National Academy of Sciences* 103: 18054-18059.
- Tang W, Newton RJ, Weidner DA (2007) Genetic transformation and gene silencing mediated by multiple copies of a transgene in eastern white pine. *Journal of Experimental Botany* 58: 545-554.
- Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.
- Wang G, Xu Y (2008) Hypocotyl-based *Agrobacterium*-mediated transformation of soybean (*Glycine max*) and application for RNA interference. *Plant Cell Report* 27: 1177–1184.
- Warthmann N, Chen H, Ossowski S, Weigel D, Hervé P (2008) Highly Specific Gene Silencing by Artificial miRNAs in Rice. *PLoS ONE* 3: e1829.
- Waterhouse PM, Graham MW, and Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proceedings of the National Academy of Sciences of the United States of America* 95: 13959-13964.
- Wesley SV, Helliwell CA, Smith NA, Wang M, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *The Plant Journal* 27: 581-590.
- Wisman E, Ramanna MS, Koornneef M (1993) Isolation of a new para mutagenic allele of the Sulfurealocus in tomato cultivar Money-maker following *in vitro* culture. *Theoretical and Applied Genetics* 87:289-294.
- Xiong AS, Yao Q-H, Peng P-H, Xian L, Han P-L, Fan, H-Q (2005) Different effect on ACC oxidase gene silencing triggered by RNA interference in transgenic tomato. *Plant Cell Reports* 23: 639-646.
- Yamada O, Ikeda R, Ohkita Y, Hayashi R, Sakamoto K, Akita O (2007) Gene Silencing by RNA Interference in the Koji Mold *Aspergillus oryzae*. *Bioscience, Biotechnology and Biochemistry* 71: 138-144.

Zuo Y, Fahey Jr GC, Merchen NR, Bajjalieh NL (1996) Digestion responses to low oligosaccharide soybean meal by ileally-cannulated dogs. *Journal of Animal Science* 74: 2441–2449.

ANEXO I

pGEM-GmSTS Glyma19g40550	MAPPNNPVNSTLGFKSLEKVF DL CDGKFTVRGVPLLSQVPNNVTFSSFSICEPRDAPPS 60 MAPPNNPVNSTLGFKSLEKVF DL CDGKFTVRGVPLLSQVPNNVTFSSFSICEPRDAPPS 60
pGEM-GmSTS Glyma19g40550	ILQRVIAVSHKGGFFGFSQVSPSDRLTNSLGSFSGRNFLSIFRFKTWWSTQVWGN SGDL 120 ILQRVIAVSHKGGFFGFSQVSPSDRLTNSLGSFSGRNFLSIFRFKTWWSTQVWGN SGDL 120
pGEM-GmSTS Glyma19g40550	QMETQWVLI E IEPEIKSYVVI I PIIEKSF R SALHPGSDGHVMICAESGSTQVKASSFGAIA 180 QMETQWVLI E IEPEIKSYVVI I PIIEKSF R SALHPGSDGHVMICAESGSTQVKASSFGAIA 180
pGEM-GmSTS Glyma19g40550	YVHVSENPYNMKEAYSVLRVHLD S FRLL E EKTVPK I ADKFGWCTWDAFYLT V NPVGVWH 240 YVHVSENPYNMKEAYSVLRVHLD S FRLL E EKTVPK I ADKFGWCTWDAFYLT V NPVGVWH 240
pGEM-GmSTS Glyma19g40550	GLKDFAE G GVAPRFVI I DDGWQSVNFDGDDPNVD A KNLVLGGEQMTARLHRFE E CDKFGS 300 GLKDFAE G GVAPRFVI I DDGWQSVNFDGDDPNVD A KNLVLGGEQMTARLHRFE E CDKFGS 300
pGEM-GmSTS Glyma19g40550	YQGLLLGPNAPSFNPKTVKELI A KGIEVERL G KLRDEAVSSGVSDLSL T EIESRIVKVK 360 YQGLLLGPNAPSFNPKTVKELI A KGIEVERL G KLRDEAVSSGVSDLSL T EIESRIVKVK 360
pGEM-GmSTS Glyma19g40550	KEIDDLFGGEGKENKELCGGCCCK E NECGGIKAFIRDLRTEFKGLDDVYVWHALCGSWG G 420 KEIDDLFGGEGKENKELCGGCCCK E NECGGIKAFIRDLRTEFKGLDDVYVWHALCGSWG G 420
pGEM-GmSTS Glyma19g40550	VRPGATHLNSKITPCKLSPGLDGTMQDLAVKIVEGSI G LVHPDQANDLYDSMHSYLAQS 480 VRPGATHLNSKITPCKLSPGLDGTMQDLAVKIVEGSI G LVHPDQANDLYDSMHSYLAQS 480
pGEM-GmSTS Glyma19g40550	GVTGVI K IDVFHSLEYVCEEYGRVELAKAYDGLTNSIVKNFNGSGIIASM Q QCNDFFFL 540 GVTGVI K IDVFHSLEYVCEEYGRVELAKAYDGLTNSIVKNFNGSGIIASM Q QCNDFFFL 540
pGEM-GmSTS Glyma19g40550	GTKQIPMGRVGDDFWFQDPNGDPMGVFWLQGVHMI R CAYNSLWMGQMIQP D WDMFQSDHV 600 GTKQIPMGRVGDDFWFQDPNGDPMGVFWLQGVHMI R CAYNSLWMGQMIQP D WDMFQSDHV 600
pGEM-GmSTS Glyma19g40550	CAK F HAGSRAICGGPVYVSDSVGSHDFDLIKMLVFPDGTVPKCIHFALPTRDCLFKNPLF 660 CAK F HAGSRAICGGPVYVSDSVGSHDFDLIKMLVFPDGTVPKCIHFALPTRDCLFKNPLF 660
pGEM-GmSTS Glyma19g40550	DQKT V LKIWNFNKYGGVIGAF N CQAGWDPKMKKIKGFSECYRPI S CTVHVTEVEWDQKK 720 DQKT V LKIWNFNKYGGVIGAF N CQAGWDPKMKKIKGFSECYRPI S CTVHVTEVEWDQKK 720
pGEM-GmSTS Glyma19g40550	EAVHMGKAE E YVYLNQAEELHFMTPKSEPLQFTIQPSTFEIYNFVPEKLGGSIKFAP I 780 EAVHMGKAE E YVYLNQAEELHFMTPKSEPLQFTIQPSTFEIYNFVPEKLGGSIKFAP I 780
pGEM-GmSTS Glyma19g40550	GLTNMFNSGGTIQELECVEKGAKVKVKG D GRFLAYSSESPKKFQLNGSDVAFEWLPD G KL 840 GLTNMFNSGGTIQELECVEKGAKVKVKG D GRFLAYSSESPKKFQLNGSDVAFEWLPD G KL 840
pGEM-GmSTS Glyma19g40550	TLNLAWIEENG V SDLAIFF 860 TLNLAWIEENG V SDLAIFF 860

Figura 1A: Comparação entre as sequências de aminoácidos deduzidas do clone pGEM-GmSTS e do gene da estaquiose sintase Glyma19g40550 disponível no Phytozome GBrowser (<http://www.phytozome.net/soybean>). O alinhamento, utilizando o programa ClustalW, mostrou uma identidade de 99% entre as sequências. As substituições não conservativas estão marcadas em cinza, as substituições conservativas em verde e os aminoácidos não marcados são idênticos.