IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES CODIFICANTES DE PROTEÍNAS RICAS EM GLICINA LIGANTES DE RNA EM SOJA (*Glycine max* (L) MERRIL)

Liane Balvedi Poersch

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES CODIFICANTES DE PROTEÍNAS RICAS EM GLICINA LIGANTES DE RNA EM SOJA (*Glycine max* (L.) MERRIL)

Liane Balvedi Poersch

Porto Alegre

2011

LIANE BALVEDI POERSCH

IDENTIFIÇÃO E CARACTERIZAÇÃO DE GENES CODIFICANTES DE PROTEÍNAS RICAS EM GLICINA LIGANTES DE RNA EM SOJA (Glycine max (l.) MERRIL)

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de mestre em Genética e Biologia Molecular

Orientadora:
Maria Helena Bodanese Zanettini
Co-Orientadora:
Márcia Maria Auxiliadora Naschenveng Pinheiro Margis

Porto Alegre

2011

AGRADECIMENTOS

Às minhas duas orientadoras Maria Helena e Márcia eu sou eternamente grata por todo o conhecimento passado, dedicação e incentivo que demonstraram durante estes dois anos de convivência. Cresci muito profissionalmente e como pessoa tendo-as como um exemplo a ser seguido. Agradeço pela oportunidade que me proporcionaram e pela querida acolhida em seu local de trabalho, também pelo enorme dedicação e esforço para disponibilizar um ambiente de trabalho onde é possível crescer e realizar todo o necessário. Muito obrigada, pois graças a vocês pude escrever com felicidade esta dissertação.

Agradeço a todos os colegas de bancada (de todos os grupos) pela ajuda e conversas que tivemos duarante o decorrer do trabalho. Principalmente pela paciência para responder as minhas numerosas perguntas (realmente não foram poucos os "porquês").

Em especial a toda equipe da soja, principalmente Marta, Lauro e Ciliana que estiveram mais envolvidos durante o trabalho. A Beatriz muito obrigada, por além de ser uma colega acabou se tornando quase uma orientadora não oficial, sendo responsável por boa parte do conhecimento prático e teórico que adquiri.

Ao nosso parceiro do Rio de Janeiro Gilberto, por ter grande colaboração e administrar o projeto em que este trabalho de encaixa.

Às fontes financiadoras Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Centro Brasileiro-Argentino de Biotecnologia (CBAB).

Por último, mas não menos importante a minha família. Agradeço aos meus pais (Rubia e Jorge) por sempre me incentivarem e cobrarem a construir uma carreira profissional cada vez melhor, bem como pelo auxílio prestado desde o primeiro ano da graduação. Ao meu marido Cristiano pelo enorme esforço que talvez eu não seja capaz de imaginar que ele fez ao me incentivar, mesmo sabendo que ficaríamos separados por pelo menos dois anos para que fosse possível a realização deste mestrado.

Muito obrigado a todos vocês!

INSTITUIÇÃO E FONTES FINANCIADORAS

Este trabalho foi desenvolvido nos laboratórios de (i) Genética e Biologia Molecular Vegetal e (ii) Cultura de Tecidos e Transformação Genética de Plantas do Departamento de Genética da Universidade Federal do Rio Grande do Sul.

Recebeu apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Centro Brasileiro-Argentino de Biotecnologia (CBAB).

RESUMO

A soja constitui uma das culturas mais importantes mundialmente, tanto social quanto economicamente. Consequentemente, informações moleculares sobre processos de desenvolvimento, bem como conhecimento detalhado das interações entre condições estressoras e a resposta da planta a fatores ambientais são necessários. A identificação e caracterização de genes que respondem a condições ambientais específicas constituem um passo inicial no entendimento dos processos adaptativos. Proteínas ricas em glicina (GRPs) são polipeptídeos contendo um grande número do aminoácido glicina em sua estrutura primária. Os genes codificantes de GRPs são regulados ao longo do desenvolvimento e regulados por auxina, ABA, frio, ferimentos, luz, ritmo circadiano, salinidade, seca, patógenos e encharcamento. Entretanto, há pouca informação sobre GRPs de plantas e seus papéis no desenvolvimento e resposta a estresses. As GRPs podem ser divididas em quatro classes (I, II, III, IV) de acordo com sua estrutura primária e presença de domínios característicos. A classe IV é composta por proteínas ligantes de RNA. Domínios adicionais permitem dividir a classe IV de GRPs em quatro subclasses (IVa, IVb, IVc, IVd). A subclasse IVc é representada por proteínas contendo um *cold-schock domain* (CSD) e dedos de zinco CCHC tipo retrovirais. O objetivo do presente estudo foi: (i) identificar e caracterizar os genes codificantes de classe IV de GRPs, (ii) verificar a padrão de expressão dos genes codificantes da subclasse IVc de GRPs e (iii) produzir plantas de soja transgênicas expressando o gene *AtGRP2*, o qual foi mostrado estar envolvido na floração e desenvolvimento da semente em Arabidopsis, e também poderia desempenhar um papel na aclimatação ao frio. Um total de 47 genes codificantes da classe IV de GRPs foi identificado no genoma da soja: 19 da subclasse IVa, sete da IVb, seis da IVc e 15 da IVd. Análises *in silico* indicaram uma expressão preferencial de todos os genes codificantes da subclasse IVc em tecidos em desenvolvimento. Análises de RT-qPCR revelaram que plantas jovens e maduras exibem uma expressão mais alta em folhas do que em outros órgãos, com exceção dos genes *GRP2L_4/5* que tiveram expressão mais alta em sementes. *GRP2L_4/5* e *GRP2L_2* foram induzidos em resposta a baixas temperaturas. Sob estresse com ABA a expressão de todos os genes foi reprimida em folhas e/ou raízes, com exceção do gene *GRP2L_2* que foi induzido em raízes. Em resposta a infecção com *Phakopsora pachyrhizi*, a expressão de *GRP2L_2* e *GRP2L_3* foi mais alta e precoce no genótipo suscetível quando comparada com o resistente, enquanto que a resposta de *GRP2L_4/5* e *GRP2L_6* foi mais tardia no genótipo resistente. Ainda, embriões somáticos secundários das cultivares Bragg, IAS-5 e BRSMG 68 Vencedora de soja foram usados para introduzir o gene *AtGRP2* no genoma da soja por bombardeamento e sistema bombardeamento/*Agrobacterium*. Seis eventos de transformação independentes foram confirmados por PCR. No presente momento as plantas estão em desenvolvimento em frascos de vidro. No presente estudo a classe IV de GRPs em soja foi identificada e caracterizada. Este é o primeiro passo para elucidar o papel destas proteínas em plantas.

Palavras-chave: Estresse abiótico. Estresse biótico. Padrão de expressão. Proteínas ricas em glicina. *Cold-shock domain*. Transformação de soja. Proteínas ligantes de RNA.

ABSTRACT

Molecular information on plant developmental process, as well as detailed knowledge of the interaction between stress conditions and plant response to environmental factors are essential for understanding the adaptive response. **G**lycine-**R**ich **P**roteins (GRP) have the amino acid glycine well represented in their primary structure. The genes encoding GRPs are developmentally regulated and induced by auxin, ABA, cold, wound, light, circadian rhythm, salinity, drought, pathogens, and flooding. However, there is scarce information about plant GRPs and its role on development and stress response. The GRPs can be divided into four classes (I, II, II and IV) according to their primary structure and the presence of characteristic domains. Class IV is composed by RNA-binding proteins. Additional domains permit to split class IV GRPs into four subclasses (IVa, IVb, IVc and IVd). Subclass IVc is represented by proteins containing a **C**old-**S**hock **D**omain (CSD) and retroviral-like CCHC zinc fingers. The goal of the present study was: (i) to identify and characterize the genes encoding class IV GRPs, (ii) to verify the relative expression of genes encoding subclass IVc GRPs and (iii) to produce transgenic soybean plants expressing the *AtGRP2* gene, which was shown to be involved in Arabidopsis flower and seed development, and can also play a role in cold acclimation. A total of 47 genes encoding class IV GRPs were found in the soybean genome: 19 from IVa, seven from IVb, six from IVc and 15 from IVd subclasses. *In silico* analyses indicated a preferential expression of all genes encoding subclass IVc GRPs in tissues under development. RT-qPCR analyses revealed that both young and mature plants exhibit relative higher expression of subclass IVc GRPs in leaves than in other organs, with exception of *GRP2L_4/5* genes that have higher expression in seeds. The *GRP2L_4/5* and *GRP2L_2* were up-regulated in response to low temperatures. Under ABA stress the expression of all genes was down-regulated in leaves and roots, with exception of *GRP2L_2* gene that was up-regulated in roots. In response to *Phakopsora pachyrhizi* infection, *GRP2L_2* and *GRP2L_3* expression was higher and earlier in the susceptible genotype when compared with that of the resistant one, while *GRP2L_4/5* and *GRP2_6* respond later in the resistant genotype. Furthermore, secondary somatic embryos of Bragg, IAS-5 and BRSMG 68 Vencedora soybean cultivars were used to introduce the *AtGRP2* gene into the soybean genome by particle bombardment and bombardment/*Agrobacterium* system. Six independent Bragg transformation events were confirmed by PCR. In the present moment the plants are under development in glass flasks. In the present study the soybean class IV GRPs were identified and characterized. This is the first step to elucidate the role of these proteins in plants.

Key-words: Abiotic stress. Biotic stress. Expression pattern. Glycine-rich proteins. Cold-shock domain. *AtGRP2.* Soybean transformation. RNA-binding proteins.

LISTA DE ABREVIATURAS E SIGLAS

2,4-D: *2,4 Dichlorophenoxyacetic acid*, ácido 2,4 Diclorofenoxiacético

ABA: *Abscisic acid*, ácido abscísico

bp: *base pairs*, pares de base

CaMV: *Cauliflower Mosaic Virus*, Vírus do Mosaico da Couve-Flor

CCHC: *retroviral-like CCHC zinc*-finger, dedo de zinco CCHC tipo-retroviral

cDNA: *complementary DNA*, DNA complementar

CSD: *Cold-Shock Domain*

Cl: *chloroplast*, cloroplasto

CSP*: Cold-Shock Protein*

Ct: *cytoplasm*, citoplasma

Down: *Down-regulation*, repressão

EST: *Expressed Sequence Tag*, Etiqueta de sequências expressas

GFP: *Green Fluorescent Protein*, Proteína cerde fluorescente

GR: *Glycine-Rich*, Rico em Glicina

GRP: *Glycine-Rich Protein*, Proteína Rica em Glicina

hai: *hous after infection*, horas após infecção

hpt: *hygromycin-phosphotransferase*, higromicina-fosfotransferase

Kan: *kanamycin*, kanamicina

KDa: quiilodaltons

Mb: Mega base

M: *mitochondria*, mitocôndria

MEME: *multiple EM for motif elicitation*

MM: *molecular marker*, marcador molecular

N: *nucleus*, núcleo

ND: *non determined*, não determinado

NJ: *neighbor joining*

np: *non predicted*, não predito

NT: *untrasformed*, não transformado

OB: *oligonulceotide/oligossacaride binding*, ligante de oligonucleotídeo/oligossacarídeo

PLANTPan: *Plant Promoter Analysis Navigator*

PIG: *Particle Inflow Gun*

PCR: *Polimerase Chain Reaction*, Reação em Cadeia da Polimerase

Rif: *rifampycin*, rifampicina

S: *secretory pathway*, via secretória

SM: *streptomycin*, streptomicina

SMART: *Simple Modular Architecture Research Tool*

SP: *spectinomycin*, spectinomicina

RT-qPCR: *Reverse- Transcription quantitative PCR*

RRM: *RNA Recognition Motif*, Motivo de Reconhecimento do RNA

SE: *standart error*, erro padrão

SP: ***S**ignal **P**eptide*, peptídeo sinal

T0: Plantas trangênicas recuperadas da cultura *in vitro*

T1: Progênie das plantas transgênicas recuperadas da cultura *in vitro*

TFBS: *Transcription Factor Binding Site*, Sítio de Ligação do Fator de Transcrição

UFRGS: Universidade Federal do Rio Grande do Sul

Up: *up-regulation*, indução

UV: *ultraviolet*, ultravioleta

SUMÁRIO

INTRODUÇÃO GERAL

1 INTRODUÇÃO GERAL

1.1 SOJA

A soja, *Glycine max* (L). MERRIL, é uma espécie anual, diplóide (2n = 40) e de autofecundação da família Fabaceae. Nativa da China integra a lista das mais antigas plantas cultivadas. Constitui uma das mais importantes culturas mundiais, correspondendo a 48% do mercado de plantas oleaginosas, sendo também uma das mais relevantes fontes de proteína para a alimentação humana (Zhang et al. 2004). A maior parte de seus grãos é moída e utilizada para a produção de farinha e óleo, destinados ao consumo tanto humano quanto animal, além de ser uma importante cultura para a produção de biocombustíveis.

Apesar do seu cultivo em larga escala ter começado apenas em 1960 (Reetz et al. 2008) possui grande importância econômica para o Brasil, segundo maior produtor mundial, tendo produzido 58,4 milhões de toneladas na safra de 2006/07 e 60 milhões de toneladas na safra de 2007/08, perdendo apenas para os Estados Unidos (CNPSO 2009a; CNPSO 2009b). Em 2008 a soja gerou um movimento de US$ 17 bilhões, com uma produção estimada de 68,60 milhões de toneladas em 2009/10 (CONAB 2010). A produtividade média da soja brasileira foi de 2.823 quilogramas por hectare, sendo o estado do Mato Grosso o maior produtor. Esta cultura destaca-se no comércio do país, em 2006 representou quase sete porcento do total de produtos exportados (CNPSO 2009a; CNPSO 2009b), sendo o principal responsável pelo crescimento da economia. Atualmente o Brasil compete com os Estados Unidos pelo título de maior produtor mundial. Previsões mostram que até o ano de 2019 o Brasil terá um aumento de 40% na produção desta oleaginosa (Tollefson 2010).

Apesar de existir um cenário positivo para a cultura da soja, todas as culturas estão expostas a estresses tanto bióticos quanto abióticos, o que diminui o rendimento e a produtividade (Varshney et al. 2009). Fatores abióticos causam a perda de milhões de dólares a cada ano devido à queda no rendimento (Mahajan and Tuteja 2005). Uma das causas de maior redução na

produtividade é o ataque de doenças, sendo a ferrugem asiática da soja, causada pelo fungo *Phakopsora pachyrhizi* a maior ameaça atual. Na safra de 2002/03, quando a doença tomou maior proporção no Brasil, o prejuízo chegou a dois bilhões de dólares (Goellner et al. 2010). Alguns métodos foram empregados para tentar conter a disseminação da doença pelo país, dentre eles estão o vazio sanitário (período de ausência de cultura) com o objetivo de diminuir a quantidade de esporos viáveis, o plantio simultâneo em macrorregiões visando evitar a disseminação de esporos para regiões mais atrasadas e a utilização de cultivares com ciclo de floração precoce (Barros 2009).

Em 2004 a soja foi recomendada como planta modelo para estudos genéticos e moleculares em leguminosas de estação quente, em adição a outras duas leguminosas de estação fria já utilizadas como modelo, *Medicago truncatula* e *Lotus japonicus* (Gepts et al. 2005). A soja teve recentemente seu genoma quase completamente seqüenciado (Schmutz et al. 2010), representando grande avanço para estudos genômicos.

Em vista da grande importância econômica para o Brasil, o melhoramento genético desta cultura é de grande interesse. Neste contexto, os estudos de genômica funcional e engenharia genética, representam ferramentas importantes para estudos básicos, podendo potencialmente fornecer subsídios para os programas de melhoramento da soja. Entender os detalhes dos mecanismos moleculares de resposta a diferentes estresses pode levar a novas estratégias para minimizar os seus impactos negativos.

1.2 FLORAÇÃO EM SOJA

A floração é controlada por uma série de fatores ambientais bem como desenvolvimentais. A complexidade dessa regulação é determinada por uma intrincada rede de transdução de sinais, através da qual o meristema apical (que origina folhas e caules) se converte em um meristema floral, que passa então a produzir órgãos florais (Zik and Irish 2003).

O fotoperíodo e a temperatura são determinantes para que ocorra a floração em soja, visto que estas condições provocam mudanças qualitativas no desenvolvimento da cultura. É importante ressaltar que há determinados períodos específicos do desenvolvimento em que a planta consegue captar ou não estas diferenças ambientais (Major et al. 1975; Embrapa 2008).

Na fase juvenil, as plantas são insensíveis ao fotoperíodo. Quando aparece a primeira folha verdadeira inicia o processo de maturação, no qual a planta passa a perceber estímulos do meio ambiente e entra na fase indutiva (passagem do período vegetativo para o reprodutivo), e a partir daí segue a antese. O tempo de permanência em cada estádio é determinado pelo grau de sensibilidade termofotoperiódica de cada genótipo (Rodrigues et al. 2001).

Em ambientes com fotoperíodo constante, a temperatura possui grande importância sobre o tempo de florescimento. Temperaturas mais baixas promovem florescimento tardio, sendo que naquelas inferiores ou iguais a 10 °C o crescimento é pequeno ou nulo, pois a floração somente é induzida acima de 13 °C, estando a faixa de crescimento ótima para a cultura em torno dos 30 °C (Embrapa 2003; Embrapa 2008).

Não havendo fotoperíodo constante, em temperaturas baixas ocorre uma diminuição no número de primórdios reprodutivos e na taxa de desenvolvimento. Em dias longos, a taxa de desenvolvimento dos órgãos reprodutivos é menor (Rodrigues et al. 2001).

A soja pode ser classificada como uma planta de dia curto, mesmo possuindo ampla variabilidade genética de resposta às exigências fotoperiódicas. A maioria das cultivares é sensível a mudanças de latitudes e datas de semeadura devido às suas respostas a variações no fotoperíodo (Hartwig and Kiihl 1979). As regiões tropicais possuem um fotoperíodo mais curto, causando assim um florescimento precoce e consequentemente redução na produtividade (Shanmugasundaram and Tsou 1978).

Existem modelos matemáticos que podem prever a data de floração. Entretanto, tais modelos possuem fórmulas complexas e resultados dificultados pela ampla combinação de fatores que influenciam a floração. Em vista disto, o estudo dos genes que controlam a passagem do estádio vegetativo para o reprodutivo é de grande importância para o aprofundamento do conhecimento deste processo.

A maioria do conhecimento sobre a regulação da floração foi estabelecida nos últimos dez anos através de estudos genéticos e moleculares realizados com a planta modelo *Arabidopsis thaliana*. Análises genéticas e fisiológicas do tempo de floração em Arabidopsis possibilitaram a identificação de um grande número de genes que regulam o processo em resposta a estímulos ambientais e endógenos. O principal mecanismo que controla o tempo da floração parece ser a repressão desta, onde genes reguladores do crescimento vegetativo reprimem genes que controlam o desenvolvimento floral. Vários genes de florescimento vêm sendo identificados e caracterizados. Dentre eles está um que codifica uma proteína rica no aminoácido glicina capaz de se ligar a RNA, o *AtGRP2* (Fusaro et al. 2007).

1.3 PROTEÍNAS RICAS EM GLICINA (GRP)

Há alguns anos um grupo de Proteínas Ricas em Glicina (GRP, do inglês ***G****lycine-****R****ich* ***P****rotein*) foi caracterizado em plantas. Desde então, um grande número de informações foi produzido, revelando que os diferentes genes que as codificam possuem padrões de expressão altamente específicos e seus produtos apresentam diversas localizações subcelulares. No entanto, até o momento as funções que essas proteínas desempenham ainda não são totalmente conhecidas.

A semelhança entre todas GRPs é a presença de domínios repetidos de glicina. Contudo, há outras regiões que não estão presentes em todas GRPs, sendo que duas delas permitem a divisão destas proteínas em dois grupos principais: (i) as que contem o peptídeo sinal (SP, do inglês ***S****ignal* ***P****eptide*; sequência alvo de direcionamento para o retículo endoplasmático), presente na maioria das proteínas descritas e (ii) as que contem sequências consenso de ligação ao RNA (revisão de Sachetto-Martins et al. 2000). Exemplos de outras regiões são o motivo de reconhecimento ao RNA (RRM, do inglês ***R****NA* ***R****ecognition* ***M****otif*), domínio conservado de oleosina, domínio *cold-shock* (CSD, do inglês ***C****old-****S****hock* ***D****omain*), dedos de zinco CCHC tipo retrovirais (CCHC), terminações carboxiterminal ricas em C (cisteína), hélice anfifílica e sequências ricas em H (histidina), P (prolina) e T (treonina) (Fusaro et al. 2001).

Segundo Sachetto-Martins et al. (2000) o domínio rico em glicina nestas proteínas pode ser descrito de maneira geral por meio da fórmula (Gly)*n*-X, onde o X representa qualquer aminoácido, incluindo glicina. Estas repetições podem ter um importante papel, pois como formam alças (*loops*) (Steinert et al. 1991) conferem grande flexibilidade a estas proteínas, gerando estruturas terciárias importantes para funções de reconhecimento molecular.

Devido à ampla modulação e variação na localização subcelular das diversas GRPs, estas proteínas não representam uma única família, mas sim um grupo, não estando necessariamente relacionadas. Ou seja, um grande grupo diverso de proteínas contendo um motivo estrutural comum (Sachetto-Martins et al. 2000).

O grupo das GRPs é composto por quatro classes, divididas de acordo com suas características primárias e presença de domínios (Fig 1). As GRPs de classe I são as mais clássicas e geralmente têm seu papel atribuído a funções estruturais devido à localização na parede celular da maioria delas. Esta classe possui na porção C-terminal um domínio rico em glicina (GR - do inglês ***G****lycine-****R****ich*), podendo conter ou não a sequência SP na região N-terminal. A classe II, após o domínio GR, possui uma região rica no aminoácido cisteína (C). Proteínas classificadas como III fazem desta a mais variada classe, possuindo o menor conteúdo de glicina e as mais caracterizadas são aquelas relacionadas a domínios oleosina. A última classe (IV) representa proteínas ligantes de RNA, possuindo domínios adicionais como o RRM, CSD e CCHC (Fusaro and Sachetto-Martins 2007), os quais levam à subdivisão desta classe em quatro grupos distintos (subclasse I, II, III e IV) (Bocca et al. 2005). No presente trabalho foi adotada a nomenclatura sugerida por Mangeon et al. (2010) para estas subclasses, IVa, IVb, IVc e IVd, respectivamente, com o objetivo de evitar possíveis confusões entre classe e subclasse.

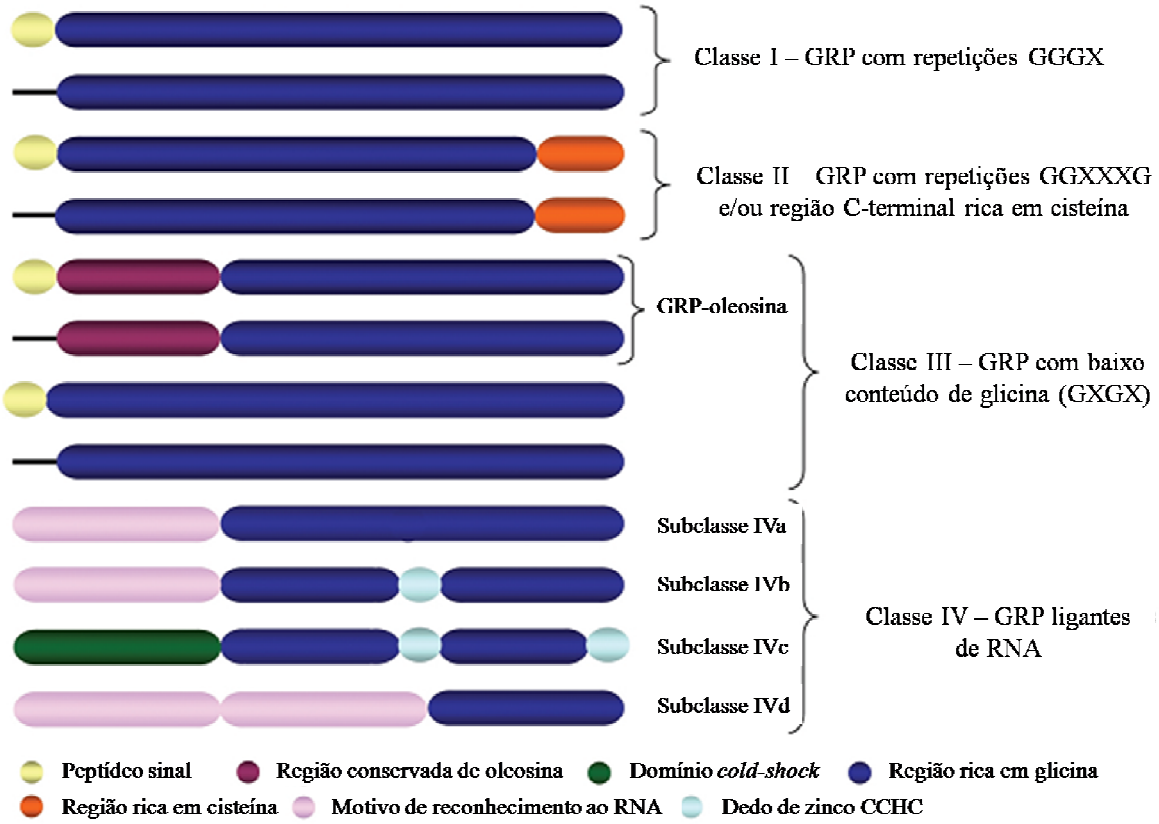

Fig 1 Representação esquemática da organização dos domínios de proteínas ricas em glicina (GRP). Fonte: (Bocca et al. 2005)

1.4 EXPRESSÃO GERAL DE GENES CODIFICANTES DE GRPs

Vários genes que codificam GRPs em plantas já foram descritos, contudo apenas poucos foram caracterizados e sua função permanece especulativa. No entanto, os dados obtidos sugerem importante papel dessas proteínas em transdução de sinais, resposta a diferentes estresses, e regulação transcricional (Bocca et al. 2005). A expressão destes genes também é regulada em função do desenvolvimento e, em alguns gêneros de plantas, através de fatores biológicos, físicos e químicos. Ainda, como estes genes apresentam um padrão bem diverso de expressão e possuem seus produtos em diferentes localizações subcelulares, é possível que estas proteínas estejam

envolvidas em muitos processos fisiológicos independentes. Pouco se sabe sobre o papel biológico dos genes codificando GRPs em plantas (Fusaro and Sachetto-Martins 2007). A compreensão das funções e importância destas proteínas em resposta a estresses é bastante limitada pela falta de dados funcionais *in vivo* (Kim et al. 2005; Kim et al. 2007).

Buscando EST (do inglês ***E**xpressed **S**equence **T**ag*) em diferentes bibliotecas (fases de desenvolvimento e diferentes estresses sofridos pelas plantas coletadas), em cana-de-açúcar, Fusaro et al. (2001) observaram a expressão diferencial tecido-específica destes genes.

Genes codificantes de GRPs são regulados por auxina, ácido abscsico (ABA, do inglês ***Ab**scisic **A**cid*), estresse hídrico, frio, ferimentos, luz, ritmo circadiano, temperatura, salinidade, seca, encharcamento e patógenos, com padrão de expressão bem diversos, variando em função do desenvolvimento (de Oliveira et al. 1990; Sachetto-Martins et al. 2000) e do tecidos/órgãos analisados. Podem ser expressos em tecidos vasculares, epiderme, flores, frutas, nódulos, tecidos em crescimento, meristemáticos e ponta de raiz, dependendo das espécies e gene em questão. Além disto, a regulação pós-transcricional vem sendo evidenciada (Sachetto-Martins et al. 2000).

Bocca et al. (2005) afirmam que os genes codificantes de GRPs são finamente regulados durante o desenvolvimento bem como sob muitas influências externas, tendo em muitos casos regulação tecido-específica. de Oliveira et al. (1990), trabalharam com cinco diferentes genes que codificam GRPs em *A. thaliana*, destes, três tem expressão preferencial em flores, sendo que todos podem ser altamente regulados em termos de especificidade de órgão e perfil desenvolvimental. Kim et al. (2005) e Kwak et al. (2005) levantam a possibilidade de que estas proteínas auxiliam na aclimatação ao frio, em temperaturas não congelantes.

1.5 GRPs LIGANTES DE RNA

As GRPs ligantes de RNA constituem um dos principias grupos das GRPs (Sachetto-Martins et al. 2000; Fusaro et al. 2007). Em plantas, tais proteínas provavelmente estão

envolvidas em respostas celulares a estímulos ambientais e desenvolvimentais, entretanto um possível papel no metabolismo do RNA permanece desconhecido (revisão de Sachetto-Martins et al. 2000).

Já foi observado que algumas GRPs ligam-se ao RNA *in vitro* e também ao DNA fita simples. Alguns estudos mostram que o domínio de ligação ao RNA e a região C-terminal rica em glicina são essenciais para esta atividade, levando à especulação que estas GRPs estejam envolvidas no processamento do RNA, na maturação e no controle da expressão gênica (Fusaro et al. 2001; Fusaro et al. 2007).

Esta classe, além dos domínios GR, é caracterizada por apresentar, na metade N-terminal da proteína, os domínios RRM ou CSD e, na porção C-terminal, pode ou não apresentar variado número de CCHC. A maioria dos RRMs possui uma região conservada de 80 a 100 aminoácidos, e dentro desta sequência duas regiões, denominadas de RNP-1 e RNP-2, são altamente conservadas (Sachetto-Martins et al. 2000). Conforme já referido, estas proteínas são divididas em quatro diferentes subclasses denominadas IVa, IVb, IVc e IVd, de acordo com os domínios presentes (Fig 1).

A subclasse IVa é caracterizada por apresentar, na metade N-terminal da proteína, um domínio RRM e, na região C-terminal, resíduos ricos em glicina com repetições RGG-box (Burd and Dreyfuss 1994; Bocca et al. 2005). Um dos membros melhor caracterizado deste grupo é a proteína MA16 de milho, a qual se localiza no núcleo e parece desempenhar um papel no metabolismo de rRNA bem como no crescimento (Gendra et al. 2004).

A subclasse IVb, em adição ao RRM na porção N-terminal e os domínios GR, possui um CCHC (Bocca et al. 2005). Um de seus representantes é o gene que codifica a proteína RZ-1 de tabaco, a qual supostamente está envolvida no processamento do pré-mRNA e/ou transporte núcleo-citoplasmático (Hanano et al. 1996).

A IVc é a subclasse das proteínas com domínio *cold-shock*. Estas proteínas não contêm o RRM, mas sim um CSD que, apesar de ser diferente, também se liga a DNA e RNA. Possuem dois ou mais CCHC dispersos dentro da região rica em glicina (Graumann and Marahiel 1998; Bocca et al. 2005). Um de seus membros mais estudado, o gene *AtGRP2* de *A. thaliana,* é descrito com maiores detalhes no item 1.6. Outra proteína integrante desta subclasse é a WCSP1 de trigo, esta

por sua vez regulada por baixas temperaturas (Karlson et al. 2002). Há indicação de que esta última proteína possua atividade de chaperona de RNA, sugerindo um possível envolvimento na aclimatação ao frio. Localiza-se no núcleo e no retículo endoplasmático, sendo que para o último é necessária a presença da porção C-terminal contendo os CCHCs (Nakaminami et al. 2006). Outro membro pertencente à subclasse IVc é NsGRP2 de *Nicotiana sylvestris* (Kingsley and Palis 1994).

As proteínas da subclasse IVd são semelhantes à IVa, possuindo domínios GR e RRM. A diferença é que enquanto na subclasse IVa há apenas um RRM, na IVd ocorrem dois ou mais RRMs (Bocca et al. 2005). Em cana de açúcar existem dois clusters representando esta subclasse, um deles usado no presente trabalho como representante desta (SCCCLR1C01G05) (Fusaro et al. 2001).

1.6 O GENE *AtGRP2*

O produto do gene *AtGRP2* possui 19 KDa, com 40% de conteúdo de glicina. O peptídeo contém um CSD na região N-terminal e um amplo domínio rico em glicina, intercalado com dois CCHC (de Oliveira et al. 1990; Kingsley and Palis 1994). Seu CSD contém somente a sequência RNP-1. Tendo em vista estas características, AtGRP2 foi incluída na subclasse IVc do grupo de GRPs ligantes de RNA (Fusaro et al. 2001). Possui localização subcelular predominantemente no citoplasma, mas também está presente no núcleo. Ao contrário, as proteínas das subclasses IVa e IVb ficam predominantemente no núcleo ou restritas ao citoplasma (Fusaro et al. 2007).

Foi observado que o gene *AtGRP2* é preferencialmente expresso em tecidos em desenvolvimento tais como meristemas, flores, óvulos, pólen, semente e primórdios foliares jovens. Plantas transgênicas de *A. thaliana* que tiveram o gene silenciado apresentaram um fenótipo de floração precoce, bem como redução do número de estames e desenvolvimento da semente afetado. Verificou-se ainda uma forte indução sob baixas temperaturas e que o mesmo não é regulado pelo ritmo circadiano (Fusaro et al. 2007).

A presença de diferentes domínios de ligação ao ácido nucléico (CSD, RGG boxes e CCHC) fortemente sugere que AtGRP2 é uma proteína ligante de RNA e/ou DNA. Em adição, domínios similares aos CCHCs foram identificados como fatores de *splicing* em leveduras e mamíferos (Fusaro et al. 2001; Fusaro et al. 2007). Além disso, foi descrito que esta proteína possui uma forte afinidade por ácido nucléico fita simples, afinidade esta que parece não ser sequência específica (Fusaro et al. 2007). Ainda não se sabe como este gene afeta a floração. Fusaro and Sachetto-Martins (2007) levantam a hipótese de que em um nível reduzido de AtGRP2 os transcritos de alguns reguladores do tempo floral tornam-se instáveis e são rapidamente degradados.

1.7 GENÔMICA

O termo genoma, definido por H. Winkler em 1920 como o material genético total de uma célula, deriva da fusão das palavras em inglês ***GEN****es* e *chromos****OME*** *– GENOME*.

A genômica é um campo da ciência que realiza estudos sistemáticos de genomas (obtenção, análise e interpretação de grandes quantidades de dados gerados por seqüenciamento do DNA) com o objetivo de entender sua estrutura, organização, função e evolução. Isto significa: localizar genes no genoma, identificar aqueles que controlam as características estudadas e analisar a expressão de milhares de genes simultaneamente. Os resultados gerados por esta área têm fornecido evidências para análises filogenéticas e funcionais (Pierce 2004).

A genômica, de uma forma geral, pode ser dividida em três grandes áreas: (i) estrutural, (ii) comparativa e (iii) funcional (Gutterson and Zhang 2004). A genômica estrutural lida com a caracterização de natureza física do genoma. Para tal baseia-se no seqüenciamento, na estrutura e organização gênica, na construção de mapas físicos e genéticos, na resolução tridimensional de proteínas, no perfil de expressão global, na genética tanto reversa quanto direta, dentre outros (Gutterson and Zhang 2004; Pierce 2004; Dita et al. 2006). A genômica comparativa realiza a comparação entre genomas de diferentes espécies, possibilitando assim associações e uma

compreensão melhor da evolução dos genomas.

O presente trabalho está inserido na terceira grande área: a genômica funcional. Esta visa determinar a função de cada parte do genoma, identificando genes, reconhecendo sua organização e entendendo suas funções. Dentre seus objetivos também estão identificar todos os RNAs transcritos em um genoma (transcriptoma) e todas as proteínas codificadas por este genoma (proteoma), fazendo uso de estudos de perfil de expressão e genética direta e reversa. Algumas técnicas utilizadas são hibridização *in situ*, experimentos de mutagênese, o uso de organismos transgênicos, perfis de expressão gênica e microarranjo (Gutterson and Zhang 2004; Pierce 2004).

A combinação do uso de técnicas da genética reversa (como exemplo super expressão ou silenciamento dos genes de interesse) com análises *high throughput* (como exemplo microarranjos e *Massively Parallel Signature Sequencing*) confere grande importância à genômica funcional, pois facilita a identificação de genes e rotas metabólicas de diversos processos biológicos (dentre eles o desenvolvimento de órgãos, a regulação do ciclo celular e a resposta a estresses ambientais).

Com todos estes dados sendo gerados tornou-se um desafio conseguir organizar e analisar claramente os resultados. É neste momento que a união com a bioinformática torna-se indispensável e de grande importância. Sem as ferramentas de bioinformática, com seus bancos de dados e programas de análises de sequências tanto de ácidos nucléicos quanto aminoácidos, em vários níveis, não seria possível a geração de conclusões sobre os dados obtidos.

1.8 TRANSFORMAÇÃO GENÉTICA DE SOJA

O desenvolvimento de protocolos de transformação possui grande importância como ferramenta para estudos genéticos, moleculares, bioquímicos e fisiológicos. Também, por meio desta técnica, genes de outras espécies bem como endógenos podem ser introduzidos, super-expressados ou silenciados, criando novos fenótipos úteis tanto para investigação da função gênica quanto para programas de melhoramento (Somers et al. 2003). Aliando-se a isto, oferece

uma alternativa para os métodos de melhoramento tradicionais de soja, que dispõe de uma estreita variabilidade genética agravada pelo fato de difíceis cruzamentos interespecíficos e intergenéricos (Hu and Zanettini 1995; Bodanese-Zanettini et al. 1996; Priolli et al. 2002).

Contudo, processos de transformação eficientes dependem de pelo menos três pré-requisitos: (i) explante adequado, (ii) metodologia de transformação eficiente e (iii) sistema de seleção de células transformadas apropriado. As plantas leguminosas apresentam um grande desafio para os processos de transformação por possuírem uma regeneração *in vitro* lenta e altamente dependente do genótipo (Somers et al. 2003; Dita et al. 2006).

Os processos de transformação genética de soja ainda não podem ser considerados como rotina em vários laboratórios, pois a resposta a estas técnicas depende do genótipo de cada cultivar, do tipo de explante a ser usado e da composição do meio de cultura (Bailey et al. 1993; Droste et al. 2002; Hofmann et al. 2004; Hiraga et al. 2007; Yang et al. 2009). No laboratório de Cultura de Tecidos e Transformação Genética de Plantas, do departamento de Genética da **U**niversidade **F**ederal do **R**io **G**rande do **S**ul (UFRGS), protocolos para a transformação e regeneração *in vitro* de plantas de soja foram estabelecidos (Droste et al. 2002; Wiebke-Strohm et al. 2011). Estes utilizam embriões somáticos secundários, provindos de cotilédones zigóticos imaturos, como explante.

Novos embriões somáticos, designados secundários, surgem da porção apical de embriões somáticos que se originaram do explante cotiledonar. Auxinas fortes como o **2,4-d**iclorofenoxiacético (2,4-D) são responsáveis por promoverem processos de desdiferenciação e reinício de divisão celular. Este regulador de crescimento faz com que células que possuam uma via de desenvolvimento determinada optem por uma via alternativa como, por exemplo, a formação de um embrião somático que dará origem a uma nova planta. Alto nível de 2,4-D (40 mg/L) no meio de cultura mantém os embriões na fase inicial de desenvolvimento (globular) e permite a proliferação de material que é alvo para a transformação.

A vantagem do uso de embriões somáticos secundários em processos de transformação genética constitui no fato de sua origem, uma única célula, da qual será formada a nova planta. Desta maneira é evitado que a planta regenerada seja quimérica, já que uma única célula que continha o transgene lhe deu origem (Finer 1988). Por esta razão, este é o explante de escolha para transformação de soja em diversos laboratórios (Finer and McMullen 1991; Sato et al. 1993;

Schmidt et al. 2008; Wiebke-Strohm et al. 2011).

A transformação genética tem sido obtida em várias espécies de plantas utilizando diferentes protocolos, os dois mais comumente usados na transformação de soja são o bombardeamento e o sistema *Agrobacterium* (Trick et al. 1997). O bombardeamento (Stanford 1988), constitui-se na aceleração de partículas de metal revestidas com o DNA de interesse contra tecidos intactos da cultura alvo. No núcleo das células, este DNA será liberado. Sua principal vantagem constitui-se no fato de não ser genótipo-dependente e qualquer material com capacidade de regeneração *in vitro* poder ser usado (Finer and McMullen 1991; Finer et al. 1992).

A transformação vegetal mediada pela bactéria *Agrobacterium tumefaciens* (Horsch et al. 1985) é outro método bastante utilizado. Nesta técnica, o gene de interesse substitui o gene patogênico original da bactéria, presente no T-DNA. A vantagem sobre o bombardeamento de partículas é que o gene exógeno será endereçado especificamente ao DNA da planta, sendo inserido em poucas cópias e não havendo danos maiores no tecido vegetal. Em adição às vantagens já citadas, fragmentos maiores de DNA podem ser inseridos (Gelvin 2003; Kohli et al. 2003). Contudo uma restrição ao uso da técnica existe, a limitação da susceptibilidade da planta à bactéria.

No laboratório de Cultura de Tecidos e Transformação Genética de Plantas, do departamento de Genética da UFRGS, foi estabelecido um protocolo de transferência de genes para soja, via bombardeamento, utilizando como alvo o tecido embriogênico mantido em meio semi-sólido (Droste et al. 2002). Utilizando este protocolo, foram obtidas plantas transgênicas da cultivar IAS-5, que expressam um gene modificado *cry1Ac* de *Bacillus thuringiensis* (Homrich et al. 2008a; Homrich et al. 2008b), plantas das cultivares Bragg e IAS-5, que expressam um gene que codifica uma osmotina de *Solanum nigrum* (Weber 2007), bem como plantas expressando genes codificadores de ureases de soja (Wiebke-Strohm 2010). Paralelamente, no mesmo laboratório, foi estabelecido de forma pioneira, um protocolo combinando os métodos de bombardeamento e o sistema *Agrobacterium*. Em um primeiro trabalho, foram obtidos embriões somáticos transformados, embora não tenham sido regeneradas plantas transgênicas (Droste et al. 2000). Desde então, vários experimentos foram realizados visando à otimização desse protocolo. Tal otimização foi conseguida pela identificação de uma combinação de antibióticos eficiente, com efeitos fitotóxicos mínimos sobre o tecido embriogênico, para suprimir a bactéria após seu co-cultivo com o explante (Wiebke et al. 2006). Recentemente este método foi estabelecido e

descrito em detalhes por Wiebke-Strohm et al. (2011). Podemos afirmar que hoje contamos com mais uma metodologia eficiente para a transformação genética de soja que utiliza o bombardeamento e o sistema *Agrobacterium* de forma integrada.

OBJETIVOS

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo geral que norteou o desenvolvimento deste trabalho foi identificar e caracterizar os genes codificadores de proteínas ricas em glicina ligantes de RNA presentes no genoma da soja e analisar o fenótipo de sojas super expressando o gene *AtGRP2*.

2.2 OBJETIVOS ESPECÍFICOS

a) Identificar no genoma da soja os genes que codificam proteínas ricas em glicina ligantes de RNA;
b) Determinar o número de genes da família que compõem as proteínas contendo domínio *cold-shock*;
c) Estudar o padrão de expressão destes genes nos diferentes órgãos da planta e em resposta a diferentes estresses (bióticos e abióticos);
d) Super expressar o gene *AtGRP2* de *Arabidopsis thaliana* em soja.

CAPÍTULO I

Identification and characterization of the soybean class IV GRP encoding genes

Manuscrito a ser submetido à revista *Molecular Genetics and Genomics*

Identification and characterization of the soybean class IV GRP encoding genes

Liane B. Poersch[1]; Beatriz Wiebke-Strohm[1]; Filipi M. da Silva[2], Lauro B. Neto[1]; Maria Helena Bodanese-Zanettini[1]; Gilberto Sachetto-Martins[2]✉ and Marcia Margis-Pinheiro[1]✉

[1] Laboratório de Genética Molecular Vegetal, Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), Brazil.

[2] Laboratório de Genômica Funcional e Transdução de Sinal, Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

✉ Corresponding authors: Marcia Margis-Pinheiro; marcia.margis@ufrgs.br, phone: 0055-51-33089814, fax: +55-51-33089823, address: Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, CP 15053, CEP 91501-970, Porto Alegre, RS, Brasil. Gilberto Sachetto-Martins; sachetto@biologia.ufrj.br, phone: 0055-21-25626379, fax: +55-21-25626379, address: Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, CCS Bloco A sala A2-066, CEP 21941-617, Rio de Janeiro, RJ, Brasil.

Abstract

The identification and characterization of responsive genes to specific environmental conditions is an initial step toward the understanding of plant adaptive responses. Genes encoding GRPs are developmentally regulated and induced by biotic and abiotic stresses. Class IV GRPs is composed by RNA-binding proteins and the subclass IVc is represented by proteins containing CSD and CCHC. Forty-seven genes encoding class IV GRP members were found in the soybean genome: 19 from IVa, seven from IVb, six from IVc and 15 from IVd subclasses. The subclass IVc encoding genes showed a preferential expression in tissues under development. Furthermore, both young and mature plants exhibit relative higher expression of genes of this subclass in leaves than in other organs, with the exception of *GRP2L_4/5* that have higher expression in seeds. The *GRP2L_4/5* and *GRP2L_2* were up-regulated in response to cold. Under ABA stress the expression of subclass IVc encoding genes was down-regulated in leaves and/or roots, with exception of *GRP2L_2* that was up-regulated in roots. In response to *Phakopsora pachyrhizi* infection, *GRP2L_2* and *GRP2L_3* expressions were higher and earlier in the susceptible genotype, while *GRP2L_4/5* and *GRP2_6* responded later in the resistant genotype.

Key-words: Glycine-Rich Protein. Biotic stress. Fungal infection. *Phakopsora pachyrhizi*. ABA. Cold stress.

Introduction

Glycine-**R**ich **P**roteins (GRP) are polypeptides that have the amino acid glycine well represented in their primary structure. These proteins contain a domain that can be described by the common formula (Gly)*n*-X, where *n* is the number of glycine residues and X represents any amino acid, including glycine. In addition to this motif, some proteins also have other domains such as **S**ignal **P**eptide (SP), **R**NA **R**ecognition **M**otif (RRM), oleosin domains, **C**old-**S**hock **D**omain (CSD), retroviral-like CCHC zinc-fingers (CCHC), among others (Sachetto-Martins et al. 2000).

The RRM comprises a conserved sequence of 80-90 amino acids, found in one or more copies per protein, which is involved in post-transcriptional regulation processes. It seems to be an ancient structure that has been found in a diversity of kingdoms (animals, fungi, plants and Monera). Two conserved regions were identified inside the domain: RNP1 (eight amino acids) and RNP2 (six amino acids) (Albà and Pagès 1998). Despite RNA binding, RNP1 and RNP2 do not discriminate their target sequences. Specificity seems to be conferred by variable amino acids within the RRM, but out of the RNPs (Burd and Dreyfuss 1994).

The CSD is a universal and very conserved domain (Sommerville 1999) that is able of binding to ssRNA, ssDNA and dsDNA (Graumann and Marahiel 1996) and is involved in translational regulation in eukaryotes. In bacteria, it acts as a RNA-chaperone (Graumann and Marahiel 1998). Sequences RNP1 and RNP2 can also be present inside CSD. The first **C**old-**S**hock **P**rotein (CSP), CSPA, was identified in *Escherichia coli* (Jones et al. 1987). The gene encoding this protein is up-regulated in low temperatures and posses a RNA-chaperone role in cold adaptation (Jiang et al. 1997). Previous studies in *Arabidopsis thaliana* (Fusaro et al. 2007) and wheat (Karlson et al. 2002) demonstrated that the AtGRP2 (At4g38680) and WCSP1, two proteins that contain CSD, play a role in cold adaptation process in plants. So far, very few plant proteins containing CSD have been studied in a functional level. Despite all the information regarding these proteins, their role and importance *in vivo* remain to be established.

GRPs can be divided into four classes according to their primary structure, arrangement of glycines and the presence of characteristic domains. Class I includes the classical GRPs group; its structural function is attributed to its cell wall localization. These proteins display a **G**lycine-**R**ich (GR) domain in the C-terminal half and may have the SP in the N-terminal half. Class II GRPs has a cysteine-rich region in addition to the GR region. Class III GRPs is the more variable group; it possesses a lower content of glycine residues if compared to the other groups. The best characterized class III proteins are the oleosin-related GRPs. The last class, class IV, is composed by RNA-binding proteins. Additional domains permit to split class IV GRPs into four subclasses (Fusaro and Sachetto-Martins 2007). These subclasses differ by the presence of domains such as RRM or CSD in the N-terminal region of the protein and CCHC within the GR domain. Subclass IVa proteins have only one RRM domain, subclass IVb members have one RRM domain and one CCHC, subclass IVc encompasses proteins which have one CSD and two or more CCHC, and proteins from subclass IVd have two or more RRM (Bocca et al. 2005).

The expression of GRP encoding genes is developmentally regulated, and induced by different biotic and abiotic factors. In addition, several GRP encoding genes show tissue-specific expression (Bocca et al. 2005). As reviewed by Sachetto-Martins et al. (2000), the GRP encoding genes have their expression regulated by auxin, **ab**scisic **a**cid (ABA), water stress, circadian rhythm, cold, wound, light, pathogens, organ and tissue, suggesting that they may play an important role in several cell processes.

GRPs do not form a single family, but a big diverse group with a single common motif: the GR domain. The great divergence among these proteins do not allow a consistent alignment of the different GRPs, with exception of the RNA-binding and the oleosin-related domains (Sachetto-Martins et al. 2000). The alignments previously reported were usually restricted to the GR regions and have to be considered with caution.

A set of GRPs was characterized in plants several years ago. After that, new GRP encoding genes have been identified in several species but a systematic study describing and characterizing the whole set of GRP present in different species is lacking. The present study was carried out with the attempt to clarify the putative roles of these genes/proteins in soybean. Specific objectives included: (i) to identify and characterize the class IV GRP encoding genes; and (ii) to verify the relative expression of soybean subclass IVc GRP encoding genes in response to different environmental conditions. Greater attention was given to soybean genes encoding CSD containing proteins, represented by subclass IVc GRPs, due to their putative role in cold acclimation as well as in development of many organisms.

Methods

Identification, classification and characterization of soybean class IV GRP encoding genes

The search for soybean class IV GRP encoding genes was carried out by the BlastP tool available on Phytozome (http://www.phytozome.net/). The protein sequences used as queries were previously described as subclass IV GRP in sugarcane (Fusaro et al. 2001), *Eucalyptus* (Bocca et al. 2005), Arabidopsis (Galvão and Sachetto-Martins unpublished data), rice (Jorge and Sachetto-Martins unpublished data) and other plants (Sachetto-Martins et al. 2000). Sequences found were then classified in each subclass according to protein domains.

The protein sequences, in addition to the description present at Phytozome, were analyzed for the presence of domains by SMART - **S**imple **M**odular **A**rchitecture **R**esearch **T**ool (http://smart.embl-heidelberg.de/; Schultz et al. 1998; Letunic et al. 2009), PFam database (http://pfam.sanger.ac.uk/; Finn et al. 2010) and InterPro database with InterPro Scan (http://www.ebi.ac.uk/Tools/InterProScan/).

The GRP encoding genes were also characterized by the number and type of splicing forms and number of introns (using model genes present at Phytozome), glycine content and their putative subcellular localization by

MitoProt - Prediction of mitochondrial targeting sequences (http://ihg.gsf.de/ihg/mitoprot.html; Claros and Vincens 1996), Wolf PSORT - Protein Subcellular Localization Prediction (http://wolfpsort.org/), Predotar (http://urgi.versailles.inra.fr/predotar/predotar.html; Small et al. 2004), iPSORT (http://ipsort.hgc.jp/), ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/; Emanuelsson et al. 1999) and TargetP 1.1(http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson et al. 2000).

All soybean class IV GRPs sequences and representative members from other plants (*A. thaliana*, *Nicotiana tabacum*, *Saccharum officinarum* and *Zea mays*) were used for motif prediction. The software MEME (**m**ultiple **E**M for **m**otif **e**licitation) version 4.4.0 (http://meme.sdsc.edu/meme4_4_0/cgi-bin/meme.cgi) was used for motif identification, according to the following parameters: minimum and maximum motif width was set to six and 300 amino acids respectively, with any number of motif repetitions. Maximum number of motifs detection was restricted to 10. All motifs identified were searched in InterPro database with InterPro Scan, in PFam and in SMART.

Gene nomenclature

The gene/protein identification code corresponded to the locus ID available on Phytozome, Glyma refers to species *Glycine max*, the first two numbers indicate the chromosome, and the number after the character "g" corresponds to the *loci*.

In this work we suggest the name GRP2-Like (GRP2L) for the subclass IVc GRPs identified in soybean genome since the representative query was the protein encoded by *AtGRP2* gene. The *loci* ID at Phytozome are designated as follow: Glyma04g00660/*GRP2L_1*, Glyma04g43130/*GRP2L_2*, Glyma06g11560/*GRP2L_3*, Glyma11g11290/*GRP2L_4*, Glyma12g03470/*GRP2L_5* and Glyma16g23690/*GRP2L_6*.

Phylogenetic analysis

All soybean class IV GRP sequences and representative members from other plants (*Z. mays* (MA16, subclass IVa), *N. tabacum* (RZ-1, subclass IVb), *A. thaliana* (AtGRP2, subclass IVc), and *S. officinarum* (SCCCLR1C01G05.g, subclass IVd) were used for phylogenetic analysis. Multiple sequence alignments were conducted with full length protein sequences using ClustalW tool (Thompson et al 1994) from MEGA v.4.0 (Tamura et al. 2007). The phylogenetic analysis was reconstructed by two different and independent approaches: the **N**eighbor-**J**oining (NJ) and the Bayesian methods. The NJ method was performed at MEGA v4.0 (Tamura et al. 2007). The molecular distances of the aligned sequences were calculated according to the p-distance parameter, with gaps and missing data treated as pair wise deletion. Branch points were tested for significance by bootstrapping with 1000 replications. Bayesian analysis was conducted in MrBayes 3.1.2 (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003) with the mixed amino acid substitution model + gamma + invariant sites. A run of 5,000,000 generations each with two Metropolis-coupled Monte Carlo Markov chains (MCMCMC) was carried out, starting from a random tree. Markov

chains were sampled every 100 generations and the first 25% of the trees were discarded as burn-in. The remaining ones were used to compute the majority rule consensus tree (MrBayes command all compat), the posterior probability of clades and branch lengths. The unrooted phylogenetic trees of class IV GRPs were wiewed and edited using the software FigTree v.1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Structural characterization of soybean subclass IVc GRPs

The protein sequences of subclass IVc GRP encoding genes were aligned by ClustalW tool from MEGA v.4.0 (Tamura et al. 2007) and edited in GeneDoc program (http://www.nrbsc.org/gfx/genedoc/; Nicholas et al. 1997). The identity of proteins was calculated manually with the alignment carried out on ClustalW (http://align.genome.jp/).

In silico *expression pattern*

The GRP encoding genes from subclass IVc had their *in silico* expression pattern individually analyzed in the Laboratório de Genética e Expressão (LGE) – Soybean database (http://bioinfo03.ibi.unicamp.br/soja/). A search in the database using the tool "EST – Gene Expression", including all libraries available was carried out using the full length transcript sequence of each subclass IVc GRP encoding gene. The analysis was also performed using the tool "Gene Expression by SuperSAGE", in the three available libraries: (i) soybean control situation X soybean infected with Asian rust (leave) PI561356 (resistant); (ii) soybean control situation X soybean subjected to drought (root) BR 16 cultivar (more susceptible); (iii) soybean control situation X soybean subjected to drought (root) Embrapa 48 cultivar (less susceptible). The parameters for search was P-value cut-off 0,05 and mismatches 0.

Experimental expression pattern analyses

The expression pattern was experimentally verified through **Q**uantitative (**R**eal **T**ime) **P**olymerase **C**hain **R**eaction (RT-qPCR). To each gene of subclass IVc a primer pair was designed using Primer3 v0.4.0 (http://frodo.wi.mit.edu/primer3/). These primer pairs amplified fragments between 118 **b**ase **p**airs (bp) and 232 bp (details about all primers are given in Table S1).

Because of their high sequence identity, the expression of *GRP2L_4* and *GRP2L_5* genes was measured by using a primer pair that amplifies both genes equal and simultaneously, including all alternative transcripts. *GRP2L_1* was not successfully amplified in a specific way and did not produce a single fragment.

The RT-qPCR reactions were conducted in the StepOne Applied Biosystems Real Time Cycler, using the StepOne Plus Software v2.0. The cycles and reactions were carried out as follow: five minutes at 94 °C, followed by 40 cycles for 15 seconds at 94 °C, 10 seconds at 60 °C, 15 seconds at 72 °C, and a final step of a melting curve when the

samples were heated at 95 °C for 15 seconds and then heated from 60 °C to 95 °C (with intervals of temperature of 0,4 °C). The final volume of reaction was 25 µL, comprised by 12.5 µL of diluted complementary DNA (cDNA) (1:100) and 12.5 µL reaction mix (2.5 µL PCR buffer 10X, 1.5 µL $MgCl_2$ 50 mM, 0.06 µL dNTP mixture10 mM, 0.25 µL primer forward 10 mM, 0.25 µL primer reverse 10 mM, 5.37 µL ultra-pure water, 2.5 µL Syber Green (1:100000 – Molecular Probes Inc., Eugene, USA) and 0.06 µL *Platinum* Taq DNA Polimerase® (Invitrogen, São Paulo, Brazil)).

The expression data analyses were performed after comparative quantification of amplified products through $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). Each experiment (described in detail bellow) was carried out with four biological samples and technical quadruplicate, with exception of soybean rust data that had three biological samples. No-template reactions were used as negative controls. The reference genes for each experimental condition were chosen according to the programs: geNorm v3.5 (http://medgen.ugent.be/~jvdesomp/genorm/) and NormFinder (http://www.mdl.dk/publicationsnormfinder.htm). Some reference genes were previously identified for expression normalization in soybean and were analyzed: ACT11, CYP2, ELF1B (Jian et al. 2008), F-box protein family, metallo protease and CDPK-related protein quinase (Libault et al. 2008). The two best reference genes for each experimental condition were selected.

RNA extraction and cDNA synthesis

Samples of soybean tissue were collected, immediately frozen in liquid nitrogen and stored at -80 ºC. Total RNA was extracted with TRIzol® reagent (Invitrogen, Carlsbad, USA), according to the manufacturer's instructions. The RNA was diluted in 20 µL with ultra-pure water and quantified by Qubit™ System (Quant-iT™ RNA Assay Kit, Invitrogen). Approximately two µg of RNA of each sample was treated with DNase I (Promega, Madison, USA), according to the manufacturer's instruction.

The synthesis of first-strand cDNA was carried out according to the manufacturer's instructions of M-MLV Reverse Transcriptase System TM (Invitrogen, Carlsbad, USA), using the DNA-free RNA and a polyT primer.

Plant growth and sampling of different soybean organs

Seeds of Brazilian soybean (MGBR-46 Conquista cultivar) were used. To verify the expression pattern of subclass IVc GRP encoding genes in young plants, seeds were sowed in vermiculite and plants were grown for two weeks at 26 °C 16/8 hours light/dark at light intensity of 22.5 $\mu Em^{-2}s^{-1}$. Leaves and roots were separately collected.

Another set of seeds were planted in pots containing organic soil and plants were grown until complete development in a green house at 28±1 ºC with 16/8 hours light/dark at light intensity of 22.5 $\mu Em^{-2}s^{-1}$. Different plant organs were collected in two independent stages. In the first one, leaves and flowers were collected from flowering

plants (about three-month-old plants at R1 developmental stage). The flowers were grouped into three sets according to their developmental stages. An association between soybean bud size and development stage of microsporogenesis was previously reported (Lauxen et al. 2003). Flower buds with 4-5 mm in length, presenting immature anthers, were considered not fertilized. Flower buds with 6-7 mm in length, with mature anthers but unfertilized gynoecium, were considered at fertilization period and the opened flowers were considered fertilized. After about three additional weeks, when plants achieved the R5 developmental (stage with 3-4 mm seeds), steam, pod with seeds, pod without seeds, seeds and seeds without the embryonic axis were harvested. For each organ, four biological replications were collected (each replication was represented by a pool of four different plants).

Cold treatment

Seeds of soybean (IAS-5 cultivar) were sowed in vermiculite and supplied with nutrition solution (MS/2 – half of concentration of MS salts medium (Murashige and Skoog 1962) and B5 vitamins (Gamborg et al. 1968)). The plants grew at controlled conditions (28±1 °C, 16/8 hours light/dark at light intensity of 22.5 $\mu Em^{-2}s^{-1}$). Cold treatment was applied in 28 day-old plants. Half of plants were transferred to a BOD at 4 °C and 16/8 h light/dark, and the other half part, which corresponded to the control plants, was maintained at controlled conditions.

The roots and two last trifoliolate leaves from treated and control plants were collected after five, 10 and 24 h. Four biological samples of each tissue at each time were used. Each biological replicate was represented by a pool of two plants. A second experiment of cold stress was carried out, with the same design above but with temperature treatment of 15°C.

ABA treatment

Seeds of Brazilian soybean cultivar MGBR-46 Conquista were sowed in plastic cups containing vermiculite and plants were grown for two weeks at 26±1 °C 16/8 hours light/dark at light intensity of 22.5 $\mu Em^{-2}s^{-1}$. The treatment consisted of the addition of 100 µM ABA MS/2 solution (Murashige and Skoog 1962). The control group received the MS/2 solution without ABA. Leaves and roots were collected after four hours. This experiment consisted of four biological samples of each tissue. Each biological sample was represented by a pool of two plants.

Soybean rust treatment

The plants were grown in a pot-based system (six pots/genotype and three plants/pot) in a greenhouse at 28± 1 °C with 16/8 hours light/dark at a light intensity of 22.5 $\mu Em^{-2}s^{-1}$. Two genotypes, Embrapa 48 (as susceptible standard – which develops Tan lesion (van de Mortel et al. 2007) and the PI561356 (as resistant standard – which

carries the resistance to soybean rust mapped on linkage group G (Abdelnoor R. R. personal communication) were used.

Plants were inoculated with *Phakopsora pachyrhizi* spores, which were collected in the field. Uredospores were harvested from leaves exhibiting sporulating uredia and diluted in distilled water with 0,05% Tween-20 to a final concentration of 3 x 10^5 spores/mL. This solution was sprayed onto leaves in half of the plants of each genotype. The other half part of plants was sprayed with a solution lacking uredospores. After this, plants were covered with water-misted bags for one day to promote infection and to prevent the cross-contamination. The inoculation was carried out in three-week-old plants.

The sampling was made with one trifoliolate leaf from each plant at one, 12, 24, 48, 96 and 192 **h**ours **a**fter **i**nfection (hai), frozen in liquid nitrogen and stored at – 80 °C until the moment of RNA extraction. This experiment was carried out in three biological replicate (each replicate was composed of a pool of three plants) from both genotypes, for control and treatment situations.

RT-qPCR experiments statistical analyses

To analyze the difference of relative expression among different organs of soybean plants the One-Way Anova test was applied. When necessary data were transformed using the method of weighted least squares (Ivar). Bonferroni multiple comparison test was performed to compare treatments. In the cold treatment and the soybean rust treatment the test carried out was the Multifactorial Anova, using three factors: organ, treatment and time of cold exposure; and genotype, treatment and time of soybean infection. When necessary data were transformed using the method of weighted least squares (Ivar). The Student's *t*-test was applied for ABA treatment.

In silico *promoter analysis*

The *in silico* promoter analysis was carried out at **Plant P**romoter **A**nalysis **N**avigator, PlantPAN (http://plantpan.mbc.nctu.edu.tw/gene_group/index.php; Chang et al. 2008). About 2000 bp upstream to the translation initiation codon of all subclass IVc GRP encoding genes were used in this analysis. The parameters used were: core similarity 1; matrix similarity 0,75; organism chosen Arabidopsis, Medicago, Bean, Pea and Soybean; Suport>= 90%.

The nucleotide sequences of the promoter region of *GRP2L_4* and *GRP2L_5* were aligned by ClustalW tool (Thompson et al 1994) from MEGA v.4.0 (Tamura et al. 2007) and edited in the GeneDoc program (http://www.nrbsc.org/gfx/genedoc (Nicholas et al. 1997).

Results

Identification, classification and characterization of soybean class IV GRP encoding genes

In the present study, the nomenclature suggested by Mangeon et al. (2010) class IV GRPs RNA-binding was followed. A total of 47 genes encoding class IV GRPs were identified in the soybean genome: 19 genes of subclass IVa (including five presenting alternative splicing forms and five transposable elements), seven genes of subclass IVb (three of them with alternative splicing form), six genes of subclass IVc (two with more than one splicing form) and 15 genes of subclass IVd (two of them with more than one splicing form) (Table 1). Interestingly, alternative transcripts from Glyma08g39830 encode identical proteins. The same was observed for Glyma14g02020.

In silico subcellular localization analyses for subclass IVa members showed variable destinies: **n**ucleus (N), **cyt**oplasm (Ct), **chl**oroplast (Cl), **m**itochondria (M) and **s**ecretory pathway (S). The full length proteins of this subclass varied from 135 to 696 amino acid residues, the content of glycine from 16 to 40% and the number of introns from zero to six (Table 1). Subclass IVb proteins showed a predicted nuclear localization. Their protein full length varied from 176 to 279 amino acid residues, and the content of glycine from 12 to 21%. One to six introns were identiffied in genes encoding these proteins (Table 1). Subclass IVc genes encode the smallest proteins of class IV, with predicted nuclear and cytoplasmatic localization. Proteins with the CCHC domain encoded by Glyma11g11290/*GRP2L_4* and Glyma12g03470/*GRP2L_5* were predicted as nuclear, whereas proteins without CCHC encoded by some alternative transcripts were predicted as cytoplasmatic. The full length proteins of this subclass varied from 120 to 251 amino acid residues, the content of glycine from 15 to 40% and the number of introns from zero to two. Interestingly, Glyma12g03470/*GRP2L_5* presents Ribonuclease B **OB** domain (OB – oligonucleotide/oligossacaride binding) that is smaller and superposed with the CSD (Table 1). Subclass IVd is formed by the largest proteins, which were predicted as nuclear and cytoplasmatic proteins. Their full lengths varied from 352 to 538 amino acid residues, the content of glycine from 16 to 23% and the number of introns from zero to five (Table 1).

The software MEME was used to predict the motif composition of the soybean class IV GRPs (Fig 1B). Ten putative motifs with estimaed values lower than 1.6e-086 were identified (Table 2). The conserved sequences of these motifs were searched in InterPro database, SMART and PFam, and compared with already known motifs. The analysis revealed that motifs 1, 2, 3, and 9 are related to RRM. Motifs 5 and 6 correspond to CCHC and CSD domains, respectively. No clue is available so far for the putative function of motifs 4, 7, 8 and 10.

Phylogenetic analysis

The tree generated by phylogenetic analysis with full length sequences of GRPs class IV is shown in Fig. 1A. Subclass IVa formed three independent clusters in the tree. Subclasses IVb, IVc and IVd clustered as individual groups. The exceptions were the presence of Glyma08g08050IVa in subclass IVb branch (Fig 1A branch “f”),

Glyma06g33940IVd in subclass IVc branch (Fig 1A branch "e") and Glyma15g02890IVa in subclass IVd branch (Fig 1A branch "c").

Structural characterization of soybean subclass IVc GRP encoding genes

All GRPs of subclass IVc exhibited one CSD (containing a RNP1 sequence) and two CCHC, except GRP2L_6 that showed five CCHC (Table 1 and Fig S1). The identity among soybean subclass IVc GRPs varied from 35% (between GRP2L_3 and GRP2L_6) to 97% (between GRP2L_4 and GRP2L_5) (Table 3). Indeed, GRP2L_4 and GRP2L_5 proteins are almost identical, having six different amino acid residues, from which only one represents a change to a different amino acid class. In addition, with the exception of a change in the RNP1 consensus sequence, all substitutions are localized outside of the conserved domains (Fig S2). Even the substitution inside the RNP1 probably do not change its function, because it occurred in the less conserved position within the motif (Landsman 1992).

Alternative transcripts were found for many GRPs encoding genes (Table 1). Interestingly the *GRP2L_4* and *GRP2L_5* have 18 and 20 alternative splicing forms, respectively. The primary transcripts of both genes encode a protein with the subclass IVc pattern, but in proteins encoded by the alternative transcripts, some CCHCs are lacking (Table1, Figs S3 and S4). Proteins without any CCHC have had their amino acid composition changed as well as the glycine content reduced.

In silico *expression pattern*

Results from *in silico* analyses of soybean subclass IVc GRP encoding genes are presented in Tables 4 and 5. In general, the expression of all genes was detected in libraries containing tissues of young plants and seeds. In plants submitted to drought, *GRP2L_1* and *GRP2L_4* were down-regulated in the sensitive soybean genotype (BR 16 cultivar).

Experimental expression analyses

The relative expression of subclass IVc GRP encoding genes was determined by RT-qPCR. The specificity of each primer pair (Table S1) was confirmed by the presence of a single peak in the melting curve and a single fragment with the expected size in agarose gel electrophoresis. Furthermore, all amplified fragments were confirmed by DNA sequencing (data not shown).

Gene expression patterns were determined in different soybean organs. The expression of the analyzed genes was detected in all organs with the exception of *GRP2L_6* that was not detected in flowers and seeds (without the embryonic axis) (Fig 2). Both young and mature plants presented the highest expression level for *GRP2L_2*, *GRP2L_3*

and *GRP2L_6* in leaves than in other organs. On the other hand, *GRP2L_4/5* exhibited significant higher expression in seeds.

Two experiments were carried out in order to determine whether genes encoding CSD-containing proteins have their expression regulated by cold stress. Interaction among organs, treatment and time was statically significant (Fig 3). *GRP2L_2* and *GRP2L*_3 were down-regulated by cold in leaves and roots, especially after 24 hours. However, *GRP2L_2* was up-regulated at 10 hours of intense cold stress (4 ºC), returning to the basal expression level at 24 hours. On the other hand, after 24 hours of cold treatment, the *GRP2L_4/5* were up-regulated in leaves submitted to 4 and 15 ºC, as well as in roots submitted to 15 ºC. In these experiments the expression of *GRP2L_6* was not detected.

The transduction signal pathway in response to cold stress is usually induced by ABA (Shinozaki and Yamaguchi-Shinozaki 2000). After four hours of ABA treatment, the analyzed genes had their expression down-regulated in leaves and/or roots. The exception was an up-regulation of *GRP2L_2* in roots (Fig. 4).

In order to determine whether GRP encoding genes of subclass IVc are involved in defense responses against fungi, the transcript levels were measured in two different genotypes under *P. pachyrhizi* infection course. *GRP2L_2* and *GRP2L_3* (Fig. 5A and B) showed a bimodal response in the susceptible cultivar (Embrapa 48), in which higher and earlier transcript levels were detected when compared to the resistant genotype (PI 561356). Differences were statistically significant for *GRP2L_3*. However, data obtain for *GRP2L_2* could not be normalized by any appropriate method due to the great variance among means. This impaired any statistical analysis. On the other hand, *GRP2L_4/5* and *GRP2L_6* (Fig 5C and D) display higher and later expression in resistant plants when compared to susceptible plants.

In silico *promoter analysis*

The search for putative **T**ranscriptions **F**actor **B**inding **S**ites (TFBS) in promoter region of all GRP encoding genes of subclass IVc at PlantPAN followed two steps. The first one identified the TFBSs present in each individual promoter. Three TFBSs were detected in *GRP2L_1*, seven in *GRP2L_2*, 74 in *GRP2L_3*, eight in *GRP2L_4*, six in *GRP2L_5* and 63 in *GRP2L_6* promoters. The second step consisted of the identification of co-occurrence of TFBSs along all six promoter regions. This analysis revealed three common TFBSs in promoter regions of GRP encoding genes of subclass IVc (Table 6). The same TFBSs are also present in the promoter region of a known Arabidopsis GRP gene (*AtGRP2*).

Due to the high identity between *GRP2L_4* and *GRP2L_5,* a more detailed analysis was carried out in these promoter sequences. Both promoter presented ATHB1_01, ATHB5_01, RAV01, AG1, PIF3 and AGL3_01. *GRP2L_4* promoter also presented ATBH9_01 and CDC5_01. Interestingly, when all the TFBSs sites present in *GRP2L_4* and *GRP2L_5* promoter regions are compared, 20 sites containing the same TFBSs were identified (Figure 6). The

alignment of these two promoter sequences evidenced the high conservation between them, with identity of 60%, especially in the second half (Fig S5).

Discussion

Identification, classification and characterization of soybean class IV GRPs

A vast number of GRP encoding genes has been identified in different organisms. However, reports of a systematic search for these genes in whole genomes are rare. Up to now, there are only two studies describing the set of GRPs encoded, in *Eucalyptus* and sugarcane transcriptomes. In *Eucalyptus,* a total of 27 sequences encoding putative RNA-binding class IV GRPs were identified: 16, two, four and five sequences of subclasses IVa, IVb, IVc and IVd, respectively (Bocca et al. 2005). In sugarcane a total of 85 sequences encoding putative RNA-binding class IV GRPs were found, which were classified as: subclass IVa (62 sequences), subclass IVb (11), subclass IVc (10) subclass IVd (two) (Fusaro et al. 2001). In the present study a total of 47 genes encoding class IV GRPs (Table 1) were identified in the soybean genome: 19 classified as subclass IVa, seven as IVb, six as IVc and 15 as IVd. As previously observed for sugarcane and *Eucalyptus*, the subclass IVa in soybean showed higher number of genes while subclasses IVb and IVc were represented by a lower number of genes. On the other hand, the subclass IVd is larger in soybean than in other species.

The availability of organized database (Phytozome) containing detailed information about the gene sequences and predicted products allowed the identification of soybean *loci* with alternative transcripts (Table 1). However, for some *loci,* the alternative transcripts result in identical proteins, suggesting that post-transcriptional processes play important role in the regulation of these genes. The alternative transcripts for GRP encoding genes were also previously observed in other species (Hirose et al. 1993; Carpenter et al. 1994; Heintzen et al. 1994; Ruiter et al. 1997).

The soybean class IV GRPs exhibited four domains with known functions: RRM, CSD, CCHC and OB (Table 1). To be classified as class IV GRP RNA-binding, a polypeptide should display at least one RRM or CSD. These domains share a similar structure when analyzed by X-ray crystallography. Their conformation consist of a 'β-barrel structure' with two antiparallel β-sheets. The RRM consists of an arrangement of β-strands and α-helices (βαββαβ) and CSD is composed by β-strands (βββββ). The RNP1 is localized in one β-sheet, while the RNP2 is present in the opposite antiparallel β-sheet. This core arrangement plays an important role in RNA-binding (Graumann and Marahiel 1996; Manival et al. 2001; Max et al. 2007).

As proteins containing RRM and CSD, proteins containing OB domain (oligosaccharide/oligonucleotide-binding fold) are also involved in the regulation of RNA translation and in RNA turnover. In the present study, the OB domain was identified only in Glyma12g03470, inside the CSD. Interestingly, the OB-fold is also formed by two antiparallel β-sheets derived from five β-strands. All these data indicated a convergent evolution of the common RNA-binding surface (Elliott and Ladomary 2010).

Phylogenetic analysis of soybean class IV GRPs

A phylogenetic analysis of GRPs is not an easy task, because the GRP group is composed by a large group of structurally diverse proteins. The alignment of different classes of GRPs faces some problems. However, for class IV GRPs, it was possible to obtain a more reliable result (Sachetto-Martins et al. 2000), probably due to the presence of highly conserved CSD, RRM and CCHC domains.

A phylogenetic analysis of soybean class IV GRPs was achieved. In the present study, subclass IVa proteins separated in three different clusters. The division of this subclaqss were previously observed also for *Eucalyptus* (Bocca et al. 2005) and sugarcane (Fusaro et al. 2001), with sequences separated into two groups. One of the three soybean subclass IVa cluster (Fig 1A branch "a") appeared as an independent branch, apart from the other sequences in the tree. These proteins are the largest soybean class IV GRPs and present one RRM in the N-terminal half, as well as three putative motifs with unknown function in its C-terminal half (Fig 1B).

The subclass IVb grouped in one branch (Fig 1A branch "f"), its members present one RRM and a confirmed CCHC. Curiously, this branch split into two clusters, from which one contains three sequences with an additional putative CCHC (Fig 1B). The sequences were considered as putative CCHC when found by MEME, but not by SMART, PFam, InterProt and Phytozome. This CCHC may have diverged sufficiently to be not detected by these programs. Phylogenetically Glyma08g08050 is more related to the subclass IVb than to the subclass IVa, indicating that this protein was originated from subclass IVb, but according to the criteria based on the structural organization of GRPs, it must be classified as a IVa member.

As expected, the subclass IVc GRPs formed a unique group (Fig 1A branch "e"), probably due to their typical domains: CSD and at least two CCHC. Curiously, the Glyma11g11290IVc/*GRP2L_4* and Glyma12g03470IVc/*GRP2L_5* are phylogenetic closely related. In addition, these genes presented very similar promoter regions. They share practically the same TFBSs. Together these results reinforced the idea that these genes duplicated recently.

The proteins of subclass IVd, containing two RRMs, were grouped in a unique cluster, which split into two branches (Fig 1A branch "c" and "d"). The first one (Fig 1A branch "c") are formed by shorter proteins containing an unknown motif "7" (Fig 1B), while the second (Fig 1A branch "d") are formed by larger proteins without the unknown motif "7" (Fig 1B).

All together, the phylogenetic data showed that the structural classification represent only the current state of genes/proteins. This classification was based in the presence/absence of specific domains and did not take into account the origin and the evolutionary history of these genes. In spite of that, the classification based on the structural characteristic of GRPs should be maintained, since it reflects their functional role.

An additional gene encoding a protein containing one CSD and one CCHC was identified in soybean, however it was not included in our analyses, because it has not enough characteristics to be classified as a subclass IVc GRP encoding gene.

Functional characterization of soybean subclass IVc GRPs

The expressions of some GRP encoding genes are known to be tissue specific and developmentally regulated (Sachetto-Martins et al. 2000; Chaikam and Karlson 2008). The *in silico* analysis revealed a common feature among all six genes encoding soybean class IVc GRPs: the presence of ESTs in seeds containing globular-stage embryos (Table 4). The expression of *GRP2L_4* and *GRP2L_6* was observed exclusively in this tissue source. Interestingly, all the libraries containing ESTs of these genes were represented by young soybean tissues. The information about the developmental stages of tissues used in library L03 (leaves) was not available. Our experimental results corroborated the *in silico* data (Fig. 2), but transcripts were detected in most soybean organs analyzed. The ubiquitous expression could be related to the presence of RAV_01 TFBS in the promoter regions of all genes (Table 6) (Heinemeyer et al. 1999). The high expression observed in leaves could be explained by the presence of the ATHB1_01 TFBS (Table 6), which is activated by a transcriptional factor possibly involved in leaf development (Schliep et al. 2010).

The expression of some GRP encoding genes was shown to be regulated by low temperatures (Sachetto-Martins et al. 2000). The CSD is present exclusively in subclass IVc GRPs. Additionally, bacteria also encode proteins containing this conserved domain. In *E. coli,* a drastic change in temperature activates the expression of five from a total of nine CSPs (Blattner et al. 1997; Yamanaka et al. 1998). Two subclass IVc proteins of Arabidopsis and one of wheat were highly up-regulated in response to cold (Karlson et al. 2002; Park et al. 2009), whereas in rice, transcripts of other two genes encoding subclass IVc GRPs were transiently up-regulated in response to low temperature and rapidly returned to basal levels (Chaikam and Karlson 2008). These proteins were also able to complement the cold-sensitive phenotype of an *E. coli* strain (BX04) which lacks four endogenous CSPs (Karlson et al. 2002; Chaikam and Karlson 2008; Park et al. 2009). Our results showed that *GRP2L_4/5* was up-regulated in soybean leaves and roots under cold stress, whereas *GRP2L_2* was only temporarily up-regulated in leaves of plants submitted to extreme cold (4 °C) (Fig. 3). These data highly suggested the involvement of these proteins in cold acclimation. An unexpected down-regulation was observed for *GRP2L_*3 in both cold treatments and for *GRP2L_*2 submitted to 15 °C stress (Fig 3). Despite the presence of a CSD, these results suggested an involvement of these proteins in other cellular processes (Feng et al. 2001; Phadtare and Severinov 2005). This was previously observed for CSPs of *E. coli* and GRPs containing CSD, such as the human YB-1, and the AtGRP2 from Arabidopsis (Fusaro et al. 2007; reviewed by Chaikam and Karlson 2010).

GRPs from different classes could also be involved in plant defense against pathogens. The *NtCIG1* gene, which encodes a class II GRP in tobacco, was shown to be involved in defense against the TVCV virus (Ueki and Citovsky 2002). The *AtGRP7* (subclass IVa GRP) is related to Arabidopsis tolerance to *Pseudomonas syringae* (Fu et al. 2007). In barley, the fungal pathogens *Erysiphe graminis* and *Rhynchosporium secalis* induced *Hvgrp-2* (class II

GRP) and *Hvgrp-3* (class IV GRP) expression in both compatible and incompatible interactions (Molina et al. 1997). In soybean response to *P. pachyrhizi* infection, subclass IVc GRP encoding genes showed different expression pattern in susceptible and resistant genotypes (Fig 3). For *GRP2L_2* and *GRP2L_3* the expression in the tolerant cultivar was not as evident as in the susceptible cultivar, suggesting a negative involvement in the defense response. For *GRP2L_4/5* and *GRP2L_6* a later up-regulation was observed in resistant genotype, indicating a putative positive defense role.

ABA is considered a stress phytohormone involved in signaling routes related to several biotic and abiotic stresses (Shinozaki and Yamaguchi-Shinozaki 2000; Mahajan and Tuteja 2005; Asselbergh et al. 2008). It was observed that the expression of many GRP encoding genes was modulated by ABA in different plant species (Luo et al. 1992; Showalter et al. 1992; Dunn et al. 1996; Nicolas et al. 1997). In the present study, genes encoding soybean subclass IVc GRPs were generally down-regulated under ABA treatment (Fig. 4). The ATHB5_01 TFBS is present in all soybean *GRP2L* promoter sequences. ATHB5 transcription factor is down-regulated by ABA-signal transduction (Johannesson et al. 2003). Consequently, the observed GRPs down-regulation could be expected.

Another interesting finding was the *in silico* expression profile of the contrasting cultivars in response to drought. A down-regulating of *GRP2L_1* and *GRP2L_4* was associated to the more sensitive genotype under drought (Table 5). These results suggest that the basal expression of these genes is important for the soybean response to this kind of stress.

In the present study, the soybean class IV GRP RNA-binding encoding genes were identified. From a total of 47 genes, six encode proteins containing the CSD and were classified in subclass IVc. Altogether, *in silico* and *in vivo* data, indicated a developmental regulation, as well as the involvement of the subclass IVc proteins in plant response to different biotic and abiotic stresses. This study represents the initial step towards the understanding of the adaptive responses of plants GRP via expression.

References

Albà MM, Pagès M (1998) Plant proteins containing the RNA-recognition motif. Trends in Plant Science 3:15-21

Asselbergh B, De Vleesschauwer D, Hofte M (2008) Global switches and fine-tuning-ABA modulates plant pathogen defense. Mol Plant Microbe Interact 21:709-719

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-1462

Bocca SN, Magioli C, Mangeon A, Junqueira RM, Cardeal V, Margis R, Sachetto-Martins G (2005) Survey of glycine-rich proteins (GRPs) in the *Eucalyptus* expressed sequence tag database (ForEST). Genetics and Molecular Biology 28:608-624

Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science 265:615-621

Carpenter CD, Kreps JA, Simon AE (1994) Genes encoding glycine-rich *Arabidopsis thaliana* proteins with RNA-binding motifs are influenced by cold treatment and an endogenous circadian rhythm. Plant Physiol 104:1015-1025

Chaikam V, Karlson D (2008) Functional characterization of two cold shock domain proteins from *Oryza sativa*. Plant Cell Environ 31:995-1006

Chaikam V, Karlson DT (2010) Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. BMB Rep 43:1-8

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics 9:561

Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem 241:779-786

Dunn MA, Brown K, Lightowlers R, Hughes MA (1996) A low-temperature-responsive gene from barley encodes a protein with single-stranded nucleic acid-binding activity which is phosphorylated *in vitro*. Plant Mol Biol 30:947-959

Elliott D, Ladomary M (2010) Molecular Biology of RNA. Oxford University Press, USA

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005-1016

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978-984

Feng Y, Huang H, Liao J, Cohen SN (2001) *Escherichia coli* poly(A)-binding proteins that interact with components of degradosomes or impede RNA decay mediated by polynucleotide phosphorylase and RNase E. J Biol Chem 276:31651-31656

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The PFam protein families database. Nucleic Acids Res 38:D211-222

Fu ZQ, Guo M, Jeong BR, Tian F, Elthon TE, Cerny RL, Staiger D, Alfano JR (2007) A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. Nature 447:284-288

Fusaro A, Mangeon A, Junqueira RM, Rocha CAB, Coutinho TC, Margis R, Sachetto-Martins G (2001) Classification, expression pattern and comparative analysis of sugarcane expressed sequences tags (ESTs) encoding glycine-rich proteins (GRPs). Genetics and Molecular Biology 24:263-273

Fusaro AF, Bocca SN, Ramos RL, Barroco RM, Magioli C, Jorge VC, Coutinho TC, Rangel-Lima CM, De Rycke R, Inze D, Engler G, Sachetto-Martins G (2007) AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. Planta 225:1339-1351

Fusaro AF, Sachetto-Martins G (2007) Blooming time for plant glycine-rich proteins. Plant Signal Behav 2:386-387

Gamborg OL, Miller RA, Ojima K (1968) Nutrient requirements of suspension cultures of soybean root cells. Exp Cell Res 50:151-158

Graumann P, Marahiel MA (1996) A case of convergent evolution of nucleic acid binding modules. Bioessays 18:309-315

Graumann PL, Marahiel MA (1998) A superfamily of proteins that contain the cold-shock domain. Trends Biochem Sci 23:286-290

Heinemeyer T, Chen X, Karas H, Kel AE, Kel OV, Liebich I, Meinhardt T, Reuter I, Schacherer F, Wingender E (1999) Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. Nucleic Acids Res 27:318-322

Heintzen C, Melzer S, Fischer R, Kappeler S, Apel K, Staiger D (1994) A light- and temperature-entrained circadian clock controls expression of transcripts encoding nuclear proteins with homology to RNA-binding proteins in meristematic tissue. Plant J 5:799-813

Hirose T, Sugita M, Sugiura M (1993) cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco: occurrence of tissue-specific alternative splicing. Nucleic Acids Res 21:3981-3987

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314

Jian B, Liu B, Bi Y, Hou W, Wu C, Han T (2008) Validation of internal control for gene expression study in soybean by quantitative real-time PCR. BMC Mol Biol 9:59

Jiang W, Hou Y, Inouye M (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. J Biol Chem 272:196-202

Johannesson H, Wang Y, Hanson J, Engstrom P (2003) The Arabidopsis thaliana homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings. Plant Mol Biol 51:719-729

Jones PG, VanBogelen RA, Neidhardt FC (1987) Induction of proteins in response to low temperature in *Escherichia coli*. J Bacteriol 169:2092-2095

Karlson D, Nakaminami K, Toyomasu T, Imai R (2002) A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. J Biol Chem 277:35248-35256

Landsman D (1992) RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. Nucleic Acids Res 20:2861-2864

Lauxen MdS, Kaltchuk-Santos E, Hu C-y, Callegari-Jacques SM, Bodanese-Zanettini MH (2003) Association between Floral Bud Size and Developmental Stage in Soybean Microspores. Braz Arch Biol Technol 46:515-520

Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic Acids Res 37:D229-232

Libault M, Thibivilliers S, Bilgin DD, Radwan O, Benitez M, Clough SJ, Stacey G (2008) Identification of Four Soybean Reference Genes for Gene Expression Normalization. The Plant Genome 1:44-54

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408

Luo M, Liu JH, Mohapatra S, Hill RD, Mohapatra SS (1992) Characterization of a gene family encoding abscisic acid- and environmental stress-inducible proteins of alfalfa. J Biol Chem 267:15367-15374

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys 444:139-158

Mangeon A, Junqueira RM, Sachetto-Martins G (2010) Functional diversity of the plant glycine-rich proteins superfamily. Plant Signal Behav 5:99-104

Manival X, Ghisolfi-Nieto L, Joseph G, Bouvet P, Erard M (2001) RNA-binding strategies common to cold-shock domain- and RNA recognition motif-containing proteins. Nucleic Acids Res 29:2223-2233

Max KEA, Zeeb M, Bienert R, Balbach J, Heinemann U (2007) Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from Bacillus caldolyticus. FEBBs J 274:1265-1279

Molina A, Mena M, Carbonero P, Garcia-Olmedo F (1997) Differential expression of pathogen-responsive genes encoding two types of glycine-rich proteins in barley. Plant Mol Biol 33:803-810

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473-398

Nicholas KB, Jr. HBN, II DWD (1997) GeneDoc: Analysis and visualization of genetic variation. EMBNEWS.NEWS 4

Nicolas C, Rodriguez D, Poulsen F, Eriksen EN, Nicolas G (1997) The expression of an abscisic acid-responsive glycine-rich protein coincides with the level of seed dormancy in *Fagus sylvatica*. Plant Cell Physiol 38:1303-1310

Park SJ, Kwak KJ, Oh TR, Kim YO, Kang H (2009) Cold shock domain proteins affect seed germination and growth of *Arabidopsis thaliana* under abiotic stress conditions. Plant Cell Physiol 50:869-878

Phadtare S, Severinov K (2005) Nucleic acid melting by *Escherichia coli* CspE. Nucleic Acids Res 33:5583-5590

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574

Ruiter RK, Van Eldik GJ, Van Herpen RM, Schrauwen JA, Wullems GJ (1997) Characterization of oleosins in the pollen coat of *Brassica oleracea*. Plant Cell 9:1621-1631

Sachetto-Martins G, Franco LO, de Oliveira DE (2000) Plant glycine-rich proteins: a family or just proteins with a common motif? Biochim Biophys Acta 1492:1-14

Schliep M, Ebert B, Simon-Rosin U, Zoeller D, Fisahn J (2010) Quantitative expression analysis of selected transcription factors in pavement, basal and trichome cells of mature leaves from *Arabidopsis thaliana*. Protoplasma 241:29-36

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95:5857-5864

Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3:217-223

Showalter AM, Butt AD, Kim S (1992) Molecular details of tomato extensin and glycine-rich protein gene expression. Plant Mol Biol 19:205-215

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4:1581-1590

Sommerville J (1999) Activities of cold-shock domain proteins in translation control. Bioessays 21:319-325

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

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 Acid Research 22:4673-4680

Ueki S, Citovsky V (2002) The systemic movement of a tobamovirus is inhibited by a cadmium-ion-induced glycine-rich protein. Nat Cell Biol 4:478-486

van de Mortel M, Recknor JC, Graham MA, Nettleton D, Dittman JD, Nelson RT, Godoy CV, Abdelnoor RV, Almeida AM, Baum TJ, Whitham SA (2007) Distinct biphasic mRNA changes in response to Asian soybean rust infection. Mol Plant Microbe Interact 20:887-899

Yamanaka K, Fang L, Inouye M (1998) The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. Mol Microbiol 27:247-255

Table 1 Soybean class IV GRP encoding genes.

Subclass	Accession number (Phytozome)	Number of splicing forms	Splicing forms: Primary Transcript	Splicing forms: Alternative transcript	Splicing forms: Transposable element	Number of introns	Full length protein (aa)	Glycine content (%)	Domains	Protein subcellular localization prediction
IVa	Glyma04g01420	1			1	1	185	39	RRM	N/Ct
IVa	Glyma05g00400	2	1			4	274	22	RRM	M
				1		5	245	20	RRM	M
IVa	Glyma06g01470	2	1			1	182	39	RRM	N
					1	0	155	?	?	?
IVa	Glyma06g22150	1	1			2	603	24	RRM	M/Cl
IVa	Glyma08g08050	1	1			2	195	17	RRM	N
IVa	Glyma08g26900	1	1			4	245	16	RRM	M/Cl
IVa	Glyma08g44150	1	1			4	648	20	RRM	N
IVa	Glyma08g44170	1	1			3	664	20	RRM	N
IVa	Glyma11g12470	1			1	1	171	40	RRM	N/Ct
IVa	Glyma11g12480	1	1			1	156	26	RRM	N
IVa	Glyma11g12490	1	1			1	143	23	RRM)	N
IVa	Glyma11g12510	3		1		2	135	26	RRM	N/Ct
					2	1	170	33	RRM	N/Ct
						2	145	30	RRM	N/Ct
IVa	Glyma11g33670	2	1			5	298	17	RRM	N
				1		6	289	16	RRM	N
IVa	Glyma12g04710*	2			2	0	172	33	RRM	S
						1	160	36	RRM	Np
IVa	Glyma15g02890*	1	1			4	233	19	RRM	Np
IVa	Glyma17g08630	1	1			4	275	24	RRM	M
IVa	Glyma18g08590	1	1			3	664	20	RRM	N
IVa	Glyma18g08610	1	1			1	696	21	RRM	N
IVa	Glyma18g50150	1	1			4	244	20	RRM	N/Ct/Cl
IVb	Glyma05g24960	1	1			1	208	21	RRM, CCHC (1x)	N
IVb	Glyma08g39830	4	1			4	186	12	RRM, CCHC (1x)	N
				3[a]		5	176	12	RRM, CCHC (1x)	N
						5	176	12	RRM, CCHC (1x)	N
						6	176	12	RRM, CCHC (1x)	N
IVb	Glyma10g42320	1	1			3	279	20	RRM, CCHC (1x)	N
IVb	Glyma14g17930	4	1			5	178	14	RRM, CCHC (1x)	N
				3		5	175	14	RRM, CCHC (1x)	N
						6	177	14	RRM, CCHC (1x)	N
						5	167	14	RRM, CCHC (1x)	N
IVb	Glyma17g29080	2	1			5	178	14	RRM, CCHC (1x)	N
				1		5	175	14	RRM, CCHC (1x)	N
IVb	Glyma18g18860	1	1			5	177	13	RRM, CCHC (1x)	N
IVb	Glyma20g24730	1	1			3	279	21	RRM, CCHC (1x)	N
IVc	Glyma04g00660	1	1			0	194	38	CSD, CCHC (2x)	N
IVc	Glyma04g43130	1	1			0	170	28	CSD, CCHC (2x)	N

Table 1(continued)

Subclass	Accession number (Phytozome)	Number of splicing forms	Splicing forms: Primary Transcript	Splicing forms: Alternative transcript	Splicing forms: Transposable element	Number of introns	Full length protein (aa)	Glycine content (%)	Domains	Protein subcellular localization prediction
IVc	Glyma06g11560	1	1			0	176	29	CSD, CCHC (2x)	N
IVc	Glyma11g290	18	1	17		1	166	38	CSD, CCHC (1x)	N
						1	141	30	CSD, CCHC (1x)	N
						1	132	28	CSD, CCHC (1x)	N
						1	120	22	CSD, CCHC (1x)	N
						1	170	40	CSD, CCHC (1x)	N
						1	149	34	CSD, CCHC (1x)	N
						1	165	39	CSD, CCHC (1x)	N
						1	150	35	CSD, CCHC (1x)	N
						1	162	38	CSD, CCHC (1x)	N
						1	152	34	CSD, CCHC (1x)	N
						1	153	36	CSD, CCHC (1x)	N
						1	141	32	CSD, CCHC (1x)	N
						1	161	38	CSD, CCHC (1x)	N
						1	150	35	CSD, CCHC (1x)	N
						1	158	19	CSD	Ct
						1	177	29	CSD	Ct
						1	163	18	CSD	Ct
IVc	Glyma12g03470	20	1	19		0	200	40	CSD, CCHC (2x) OB	N
						1	152	34	CSD, CCHC (1x) OB	N
						1	141	32	CSD, CCHC (1x) OB	N
						2	146	33	CSD, CCHC (1x) OB	N
						1	184	37	CSD, CCHC (2x) OB	N
						1	147	33	CSD, CCHC (1x) OB	N
						1	140	30	CSD, CCHC (1x) OB	N
						1	150	34	CSD, CCHC (1x) OB	N
						1	145	31	CSD, CCHC (1x) OB	N
						1	150	35	CSD, CCHC (1x) OB	N
						1	129	27	CSD, CCHC (1x) OB	N
						2	139	33	CSD, CCHC (1x) OB	N
						1	176	34	CSD, CCHC (2x) OB	N
						1	153	35	CSD, CCHC (1x) OB	N
						1	153	36	CSD, CCHC (1x) OB	N
						1	158	19	CSD, OB	Ct
						1	180	27	CSD, OB	Ct
						1	172	24	CSD, OB	Ct
						1	149	15	CSD, OB	Ct
						2	158	19	CSD, OB	Ct
IVc	Glyma16g23690	1	1			0	251	28	CSD, CCHC (5x)	Ct
IVd	Glyma02g46650	1	1			3	477	17	RRM (2x)	N

Table 1(continued)

Subclass	Accession number (Phytozome)	Number of splicing forms	Splicing forms			Number of introns	Full length protein (aa)	Glycine content (%)	Domains	Protein subcellular localization prediction
			Primary Transcript	Alternative transcript	Transposable element					
IVd	Glyma02g47690	2	1			4	538	16	RRM (2x)	N
				1		4	495	17	RRM (2x)	N
IVd	Glyma05g09040	1	1			5	370	22	RRM (2x)	N
IVd	Glyma06g33940	1	1			0	444	21	RRM (2x)	N
IVd	Glyma08g43740	1	1			4	479	18	RRM (2x)	N
IVd	Glyma10g33320	1	1			2	471	23	RRM (2x)	N
IVd	Glyma13g11650	1	1			5	352	23	RRM (2x)	N/Ct
IVd	Glyma13g42480	1	1			5	364	16	RRM (2x)	N
IVd	Glyma14g00970	1	1			2	479	18	RRM (2x)	N
IVd	Glyma14g02020	2[a]	1			4	478	17	RRM (2x)	N
				1		3	478	17	RRM (2x)	N
IVd	Glyma16g07660	1	1			5	372	22	RRM (2x)	N/Ct
IVd	Glyma18g09090	1	1			4	476	17	RRM (2x)	N
IVd	Glyma19g00530	1	1			5	377	22	RRM (2x)	N/Ct
IVd	Glyma19g10300	1	1			5	374	22	RRM (2x)	N/Ct
IVd	Glyma20g34330	1	1			1	476	23	RRM (2x)	N

RRM (RNA Recognition Motif), CCHC (retroviral-like CCHC zinc-Finger), CSD (Cold-Shock Domain), OB (Ribonuclease B OB Domain), N (Nucleus), M (Mitochondria), Ct (Cytoplasm), Cl (Chloroplast), S (Secretory pathway) and np (non predicted)

[a] Identical proteins

[*] Prediction considering first metionine of ORF

Table 2 Conserved amino acid sequences of motif identified by MEME in soybean class IV GRPs

Motif nunber[a]	E value[b]	Sites[c]	Domain (related)	Conserved amino acids of motif (best possible match)
1	6.1e-956	50	RRM	EQLKKYFSKYGEITECKIMMDHNTQRPRGFCFITF
2	1.9e-437	50	-	VIMNKHELDGKQVEVKKAQPK
3	1.5e-342	50	RRM/SP	TKKCFIGGLAWWTTE
4	9.9e-301	4	?	PMGRMGGYGGFPGAPTPPFSGILPSFPGVGGVGLP GVAPHVNPAFFGRGMPVNGMGMMPGSVMDGPNM GMWSDPNMGGWGGEEPGGGKAGESSYGEEAASD HQYGEVSHDRAGWPMREKDRGSERDWSGTSERR YRDDRDQGYERDAPKEKDMGHDHEWSERRHRDD RETGRERSRDRDR
5	1.6e-276	25	CCHC	GGGRGGGGGKCYNCGEPGHFARDCPMEGG
6	8.3e-206	7	CSD	RVTGKVKWFNDQKGFGFITPDDGTDDLFVHQSQIR SDGFRSLAEGESVEFLIESEDDGRTKAVDVTGPDGA PVQ
7	1.0e-135	6	?	ELSPGPSRSPLIGYNYGLTRTSNFLNSFAQGYNMSP IGGYGVRMDGRFSPLTSGRSGFTPFGS
8	3.2e-134	4	?	RDRDRDREKDHRERDRHREDRDRYADHHRYRDRE AEHDDEWERGRSSRTHSKSRLSQEEEHHSRPRDA DYGKRRRLTSE
9	3.3e-097	4	RRM	ERDLEDEFRIFGVIRSVWVARRPPGYAFIEFDDRRD AQDAIQELDGKNGWRVELSHN
10	1.6e-086	4	?	WGRGNNPGMGNRGPVNPMRNRGGGMGGRGIMG HGGNGFGQGMGGTPPMLHPQSMMNQ

[a]The motif numbers correspond to the numbers in Fig. 1B

[b]The expected values of each motif prediction are given by the MEME program

[c]The number of the motif appeared in the 51 class IV GRP sequences presented in Fig. 1B

? Motif lacking functional annotations in the literature

Table 3 Identity (%) of proteins of soybean subclass IVc GRP and AtGRP2

	GRP2L_1	**GRP2L_2**	**GRP2L_3**	**GRP2L_4**	**GRP2L_5**	**AtGRP2**
GRP2L_1	-					58
GRP2L_2	48	-				46
GRP2L_3	50	82	-			43
GRP2L_4	68	50	49	-		66
GRP2L_5	66	51	50	97	-	66
GRP2L_6	40	37	35	45	44	43

Table 4 *In silico* expression pattern of subclass IVc GRP encoding genes by EST Gene Expression at LGE – Database*.

Gene	Lybrary	Tissue source	Alternative transcripts found in *GRP2L_4* and *GRP2L_5* genes
GRP2L_1	F05	Immature flowers of field grown plants	
	L03	Fully expanded leaves of greenhouse grown plants	
	S12	Seeds containing globular-stage embryos	
	R06	Root	
GRP2L_2	S12	Seeds containing globular-stage embryos	
GRP2L_3	S09	Whole seedling of greenhouse grown plants	
	S12	Seeds containing globular-stage embryos	
	H03	Hypocotyl and plumule, germinating seeds	
GRP2L_4	S12	Seeds containing globular-stage embryos	Alternative transcripts 1 to 9
GRP2L_5	S12	Seeds containing globular-stage embryos	Alternative transcripts 1 to 9
	B01	Vegetable buds of field grown plants	Alternative transcripts 1 to 9
	S01	Seeds germinated for three days	Alternative transcripts 1 to 9
	H05	Etiolated hypocotyl tissue of 9-10 days old seedling	Alternative transcripts 1 to 9
	SH1		Alternative transcripts 1 to 9
	R06	Germinating shoot, 24 hours germination	Alternative transcripts 1 to 9
		Root	Alternative transcripts 1 to 9
GRP2L_6	S12	Seeds containing globular-stage embryos	

*(http://bioinfo03.ibi.unicamp.br/soja/)

Table 5 *In silico* expression by analysis of SuperSAGE data at LGE – SoybeanDatabase**. Comparative SuperSAGE between control (untreated) and treated plants (plants under different stress conditions).

Gene	Soybean infected with Asian rust (PI561356 – sensitive genotype)	Soybean submitted to drought (root) (BR 16 cultivar – more susceptible)	Soybean submitted to drought (root) (Embrapa 48 cultivar – less susceptible)
GRP2L_1	No	Yes (down*)	No
GRP2L_2	Yes (up)	Yes (up)	No
GRP2L_3	No	No	No
GRP2L_4	Yes (up)	Yes (down*)	No
GRP2L_5	No	No	No
GRP2L_6	No	No	No

Yes = differences found between control and treatment situations; No = no differences found between control and treatment situations; Up = up-regulation; Down = Down-regulation
*statistically significant
**(http://bioinfo03.ibi.unicamp.br/soja/)

Table 6 The co-occurrence of TFBS found in soybean promoter region of subclass IVc GRP encoding genes.

	TFBS	*GRP2L_*					
		1	***2***	***3***	***4***	***5***	***6***
*	ATHB1_01	✓	✓	✓	✓	✓	✓
*	ATHB5_01	✓	✓	✓	✓	✓	✓
*	RAV1_01	✓	✓	✓	✓	✓	✓

* Also found in promoter region of *AtGRP2*

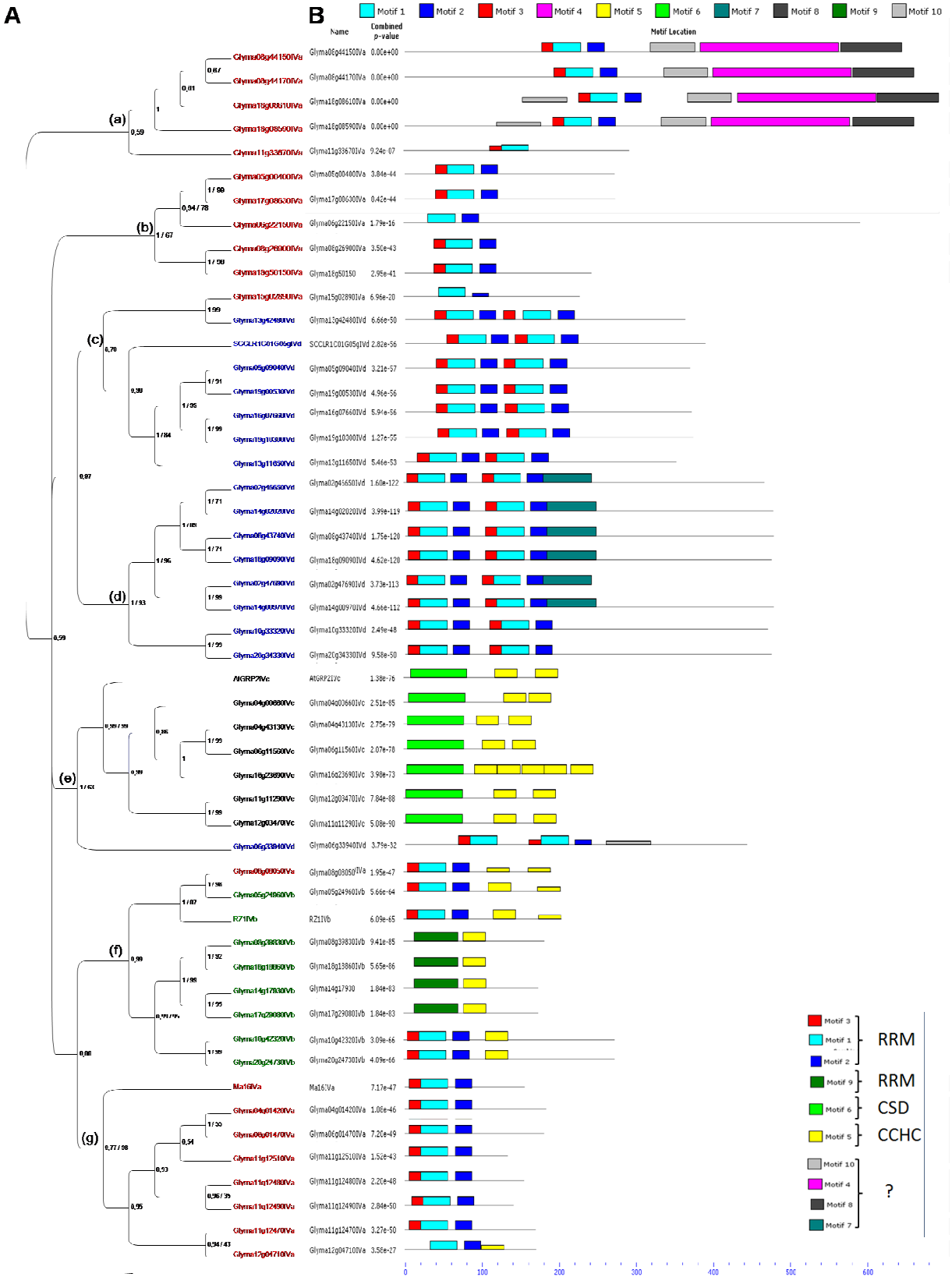
A
B
Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 6
Motif 7
Motif 8
Motif 9
Motif 10
Name
Combined p-value
Motif Location
(a)
(b)
(c)
(d)
(e)
(f)
(g)
RRM
RRM
CSD
CCHC
?
0
100
200
300
400
500
600

Fig. 1 Motif composition and phylogenetic analysis of the 51 GRPs. (A) Bayesian phylogenetic tree of 51 plant proteins class IV GRP encoding genes. The Bayesian analysis was conducted at Mr.Bayes v3.1.2, upon the alignment of GRP full length sequences protein from selected plant species at ClustalW. The unrooted cladogram was edited using Fig Tree v1.3.1. Nodal support is given by posterior probability shown next to the corresponding nodes. The bootstrap values generated by neighbor-joining analysis at MEGA were also included and are shown next posterior probabilities (separated by a slash). The scale bar indicates the estimated number of amino acids substitutions per site. The names of the soybean GRP encoding genes were identified by their *loci* ID at Phytozome plus the subclass to which it belongs (IVa in red, IVb in green, IVc in black and IVd in blue). *A. thaliana* (AtGRP2); *G. max* (Glyma); *Nicotiana tabacum* (RZ1), *Saccharum officinarum* (SCCRL1C01G05g). (B) The motif composition of GRPs was identified by MEME program. Numbered color boxes represent different putative motifs. The gene names and *loci* IDs (plus the subclass) are in accordance with the nomenclature described in Fig. 1A. The expected values calculated by MEME are shown after the gene names/*loci* IDs. Combined block diagrams indicate non-overlapping sites with a *p*-value better than 0.0001. The height of the motif "block" is proportional to –log(p-value), truncated at the height for a motif with a p-value of 1e-10.

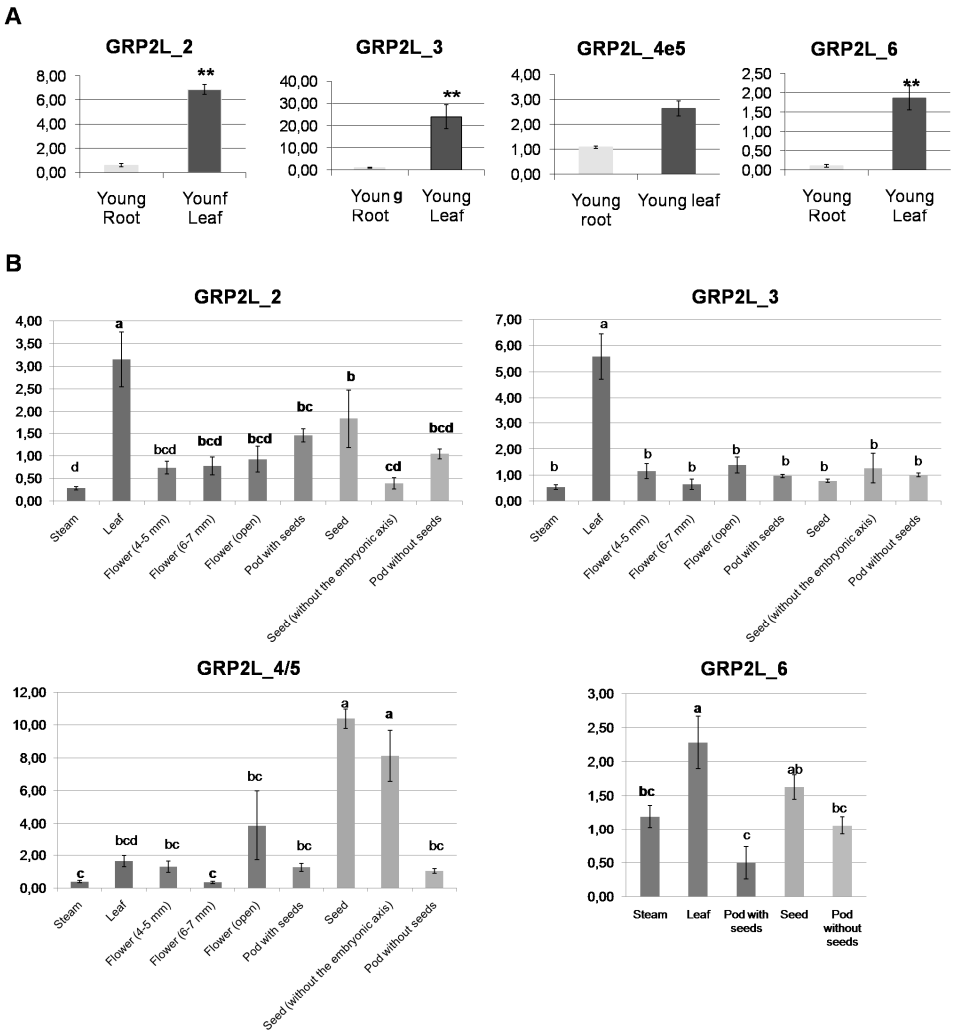

Fig. 2 RT-qPCR relative expression of subclass IVc GRP encoding genes in different organs and stages of development of soybean plants. (A) Expression in young plants (two-week-old). Data represent means of four biological replicates (±SE) and four technical replicates. "**" Means statistically different (Student's *t*-test, $p \leq 0{,}01$). (B) Expression in mature plants. Data represent means of four biological replicates (±SE) and four technical replicates. Statistical test: One-Way ANOVA; the letters represent the Bonferroni multiple comparison test, means followed by equal letters do not differ significantly. The Y axis of graphics are relative expression. F-box protein and metalloprotease encoding genes were used as reference genes.

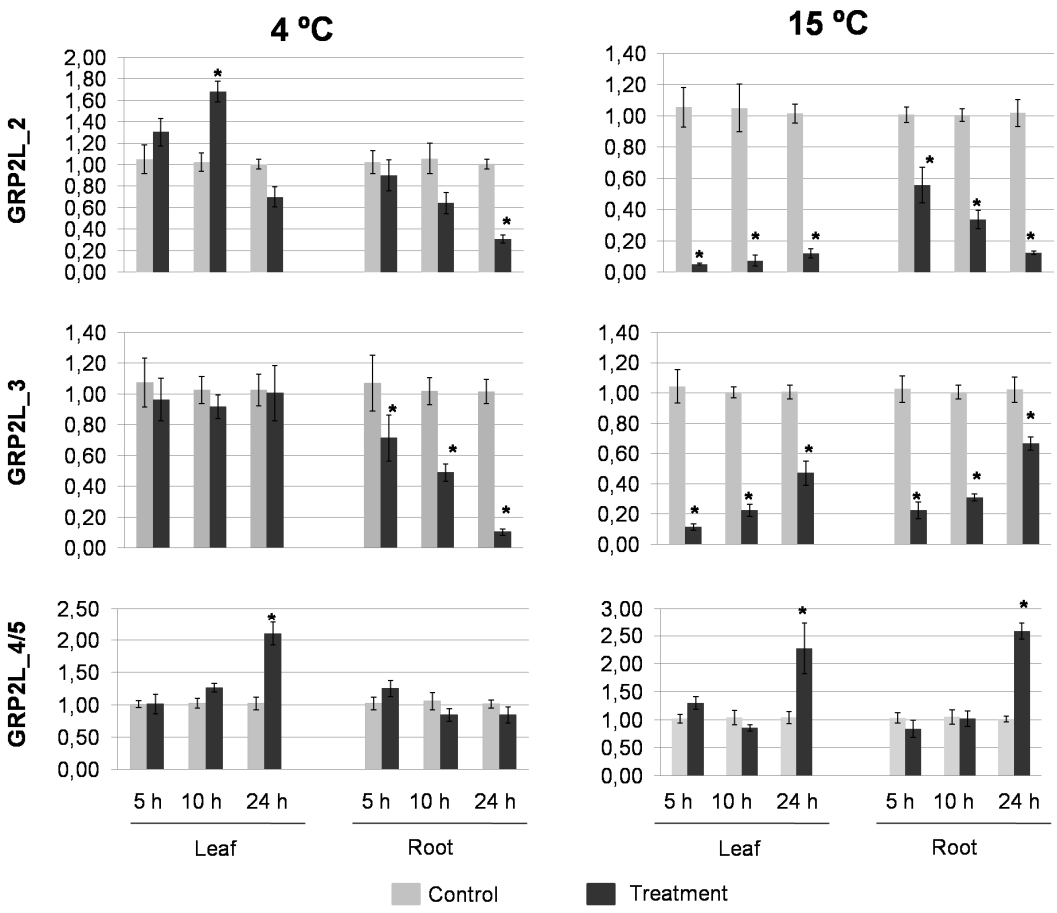

Fig. 3 RT-qPCR relative expression of subclass IVc GRP encoding genes in response to cold-shock treatment. The Y axis of graphics represents the relative expression. Data correspond to means of four biological replicates (±SE) and four technical replicates. Statistical analysis: Multifactorial ANOVA (organ/treatment/time). F-box protein and metalloprotease encoding genes were used as reference genes to 4 ºC treatment. ACT 11 and CDPK-related protein quinase encoding genes were used as reference genes to 15ºC treatment.

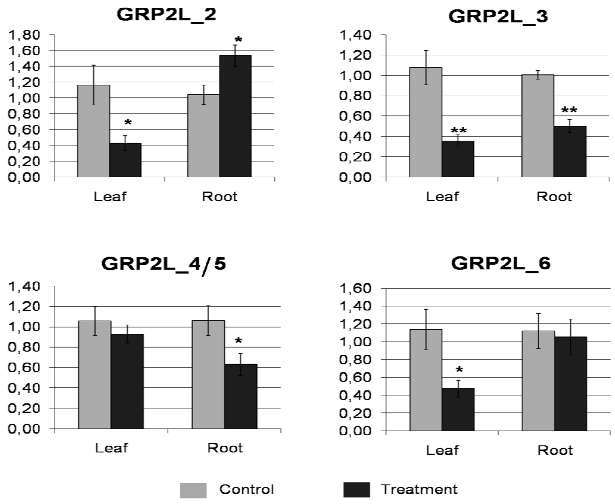

Fig. 4 RT-qPCR relative expression data of GRP encoding genes in response to four hours after 100 µM ABA treatment. The Y axis of graphics represents the relative expression. Data are means of four biological replicates (±SE) and four technical replicates. "*" Means are statistically different (Student's *t*-test, $p \leq 0{,}05$); "**" Means are statistically different (Student's *t*-test, $p \leq 0{,}01$). The reference genes used were F-box protein and CYP2.

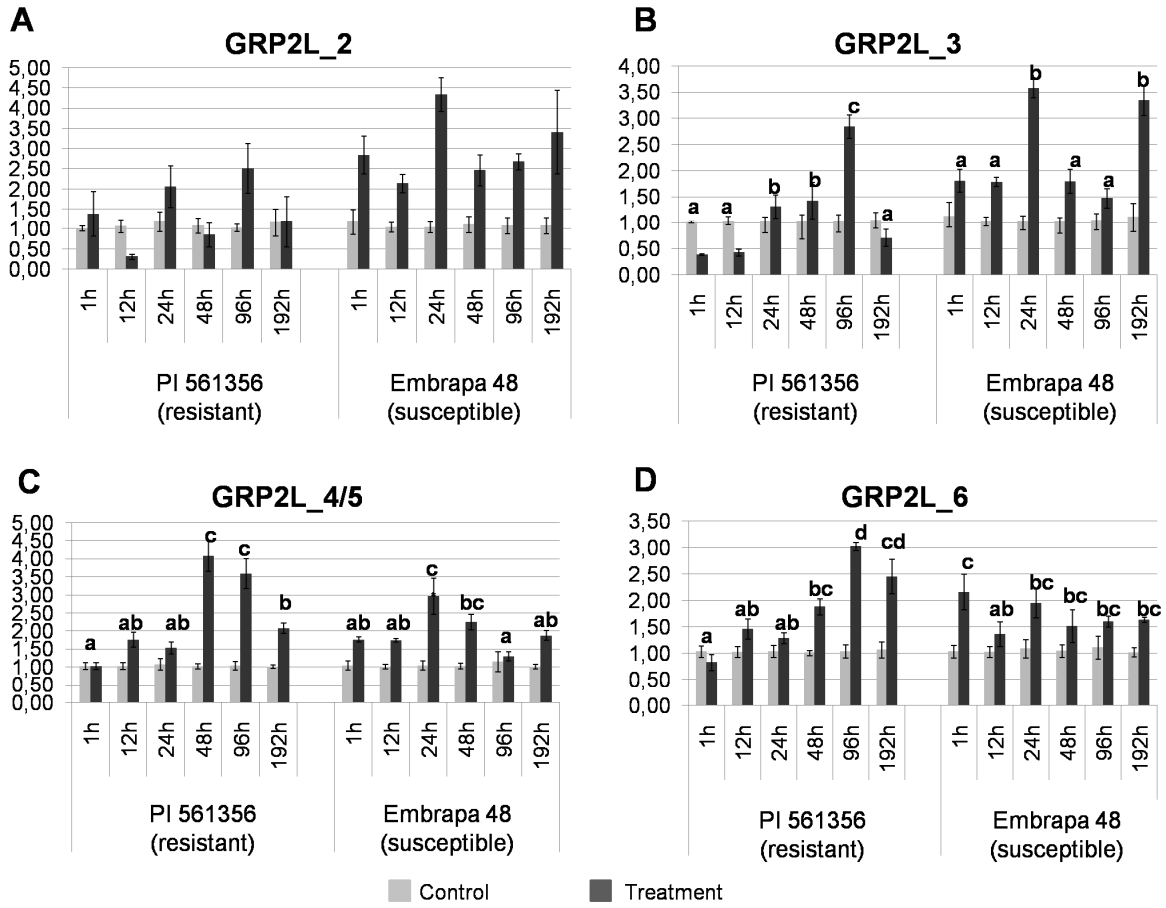

Fig. 5 RT-qPCR relative expression data of soybean GRP encoding genes in response to the *Phakopsora pachyrhizi* infection. The Y axis of graphics represents the relative expression.
Values are means of three biological replicates (±SE) and four technical replicates. Statistic test: Multifactorial Anova (treatment/genotype/time). F-box protein and metalloprotease encoding genes were used as reference genes.

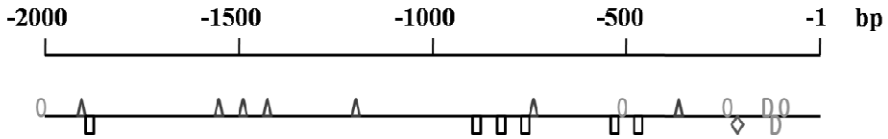

Fig. 6 Schematic co-occurrence of transcription factor binding sites of *GRP2L_4* and *GRP2L_5* promoter regions . "○" RAV1_01; "Δ" ATBH1_01; "□" ATBH5_01; "◊" AGL3_01; "D" PIF3_01.

Table S1 Primer set designed for RT-qPCR

Target	Orientation	TM (°C)	Primer sequence	PCR product size (bp)
GRP2L_2	Forward	60,53	5'AGGAAATGAGTGAGAATGTGTTATGTT3'	163
	Reverse	59,31	5'AAACAATTTAGCAAAGCGTGG3'	
GRP2L_3	Forward	59,28	5'GGAATGTCCGAATGTTGGA3'	155
	Reverse	59,74	5'AACATATTTGGAAGGAAGGAAGG3'	
GRP2L_4/5	Forward	60,12	5'TCTTAGGGATTTTAATGGTTTCTGTC3'	118
	Reverse	59,60	5'TTTGCAGATTGTAACGATAACACA3'	
GRP2L_6	Forward	61,91	5'AATGTGGGTGTAGGCGGC3'	232
	Reverse	62,19	5'ACGTGCTTTTTCCACCGC3'	

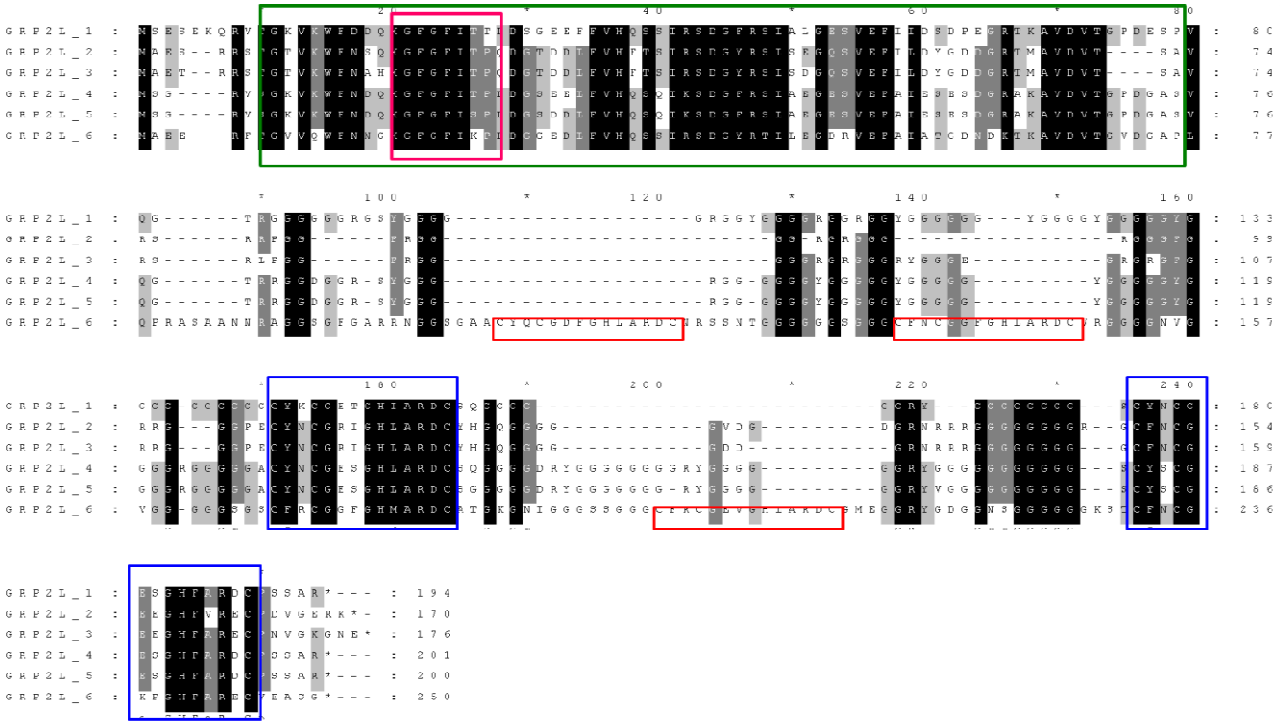

Fig. S1 Alignment of subclass IVc GRPs showing the domains. CSD: Green box, RNP1: Pink box, CCHC present in all sequences: Blue boxes, CCHC present only in GRP2L_6: Red boxes. The black–gray scale background show the conservation of amino acids among sequences, the more darkness more conserved (conserved percent in black is 100, in dark gray is 80, is light gray 60).

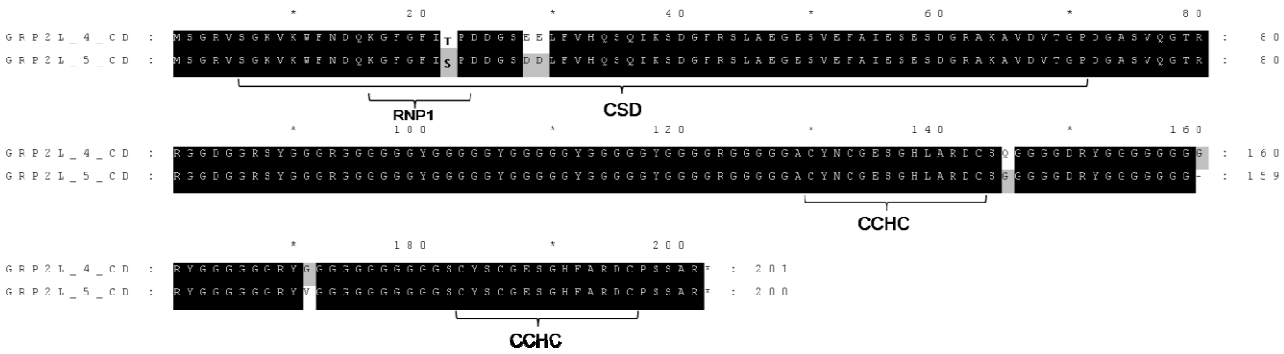

Fig. S2 Protein alignment of GRP2L_4 and GRP2L_*5*. The black background represents identical amino acids. The different amino acids are shown by the gray and white backgrounds.

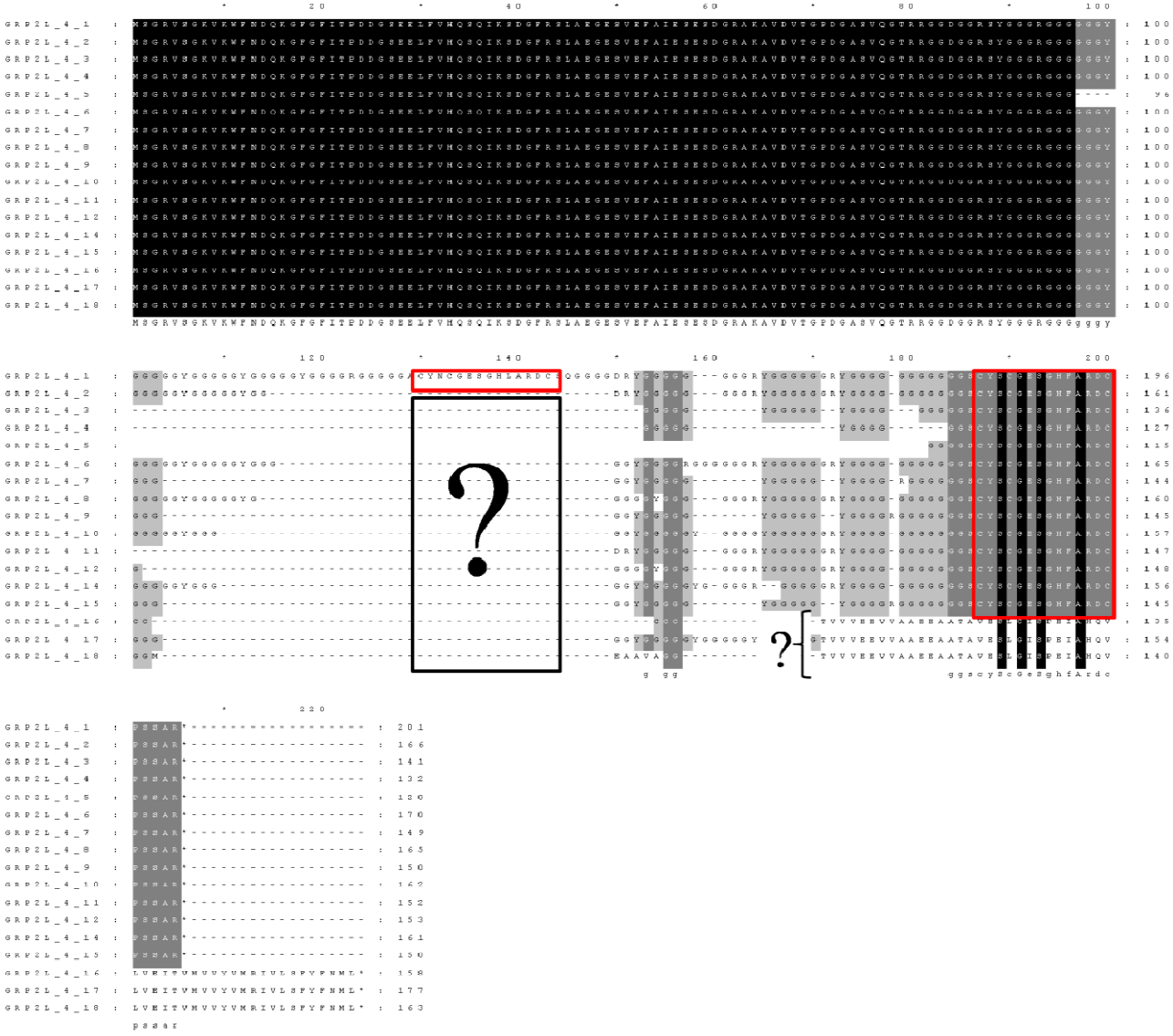

Fig. S3 Alignment of amino acids sequences predicted form the alternative transcript of *GRP2L_4*. The black–gray scale background shows the conservation of amino acids among sequences, the more darkness more conserved (conserved percent in black is 100%, in dark gray is 80% and in light gray is 60%). In the first row is presented the CSD, red boxes indicates the presence of CCHC and the "?" indicates the lack of CCHC. The number of alternative transcripts accordingly to Phytozome is after the gene name *GRP2L_4*.

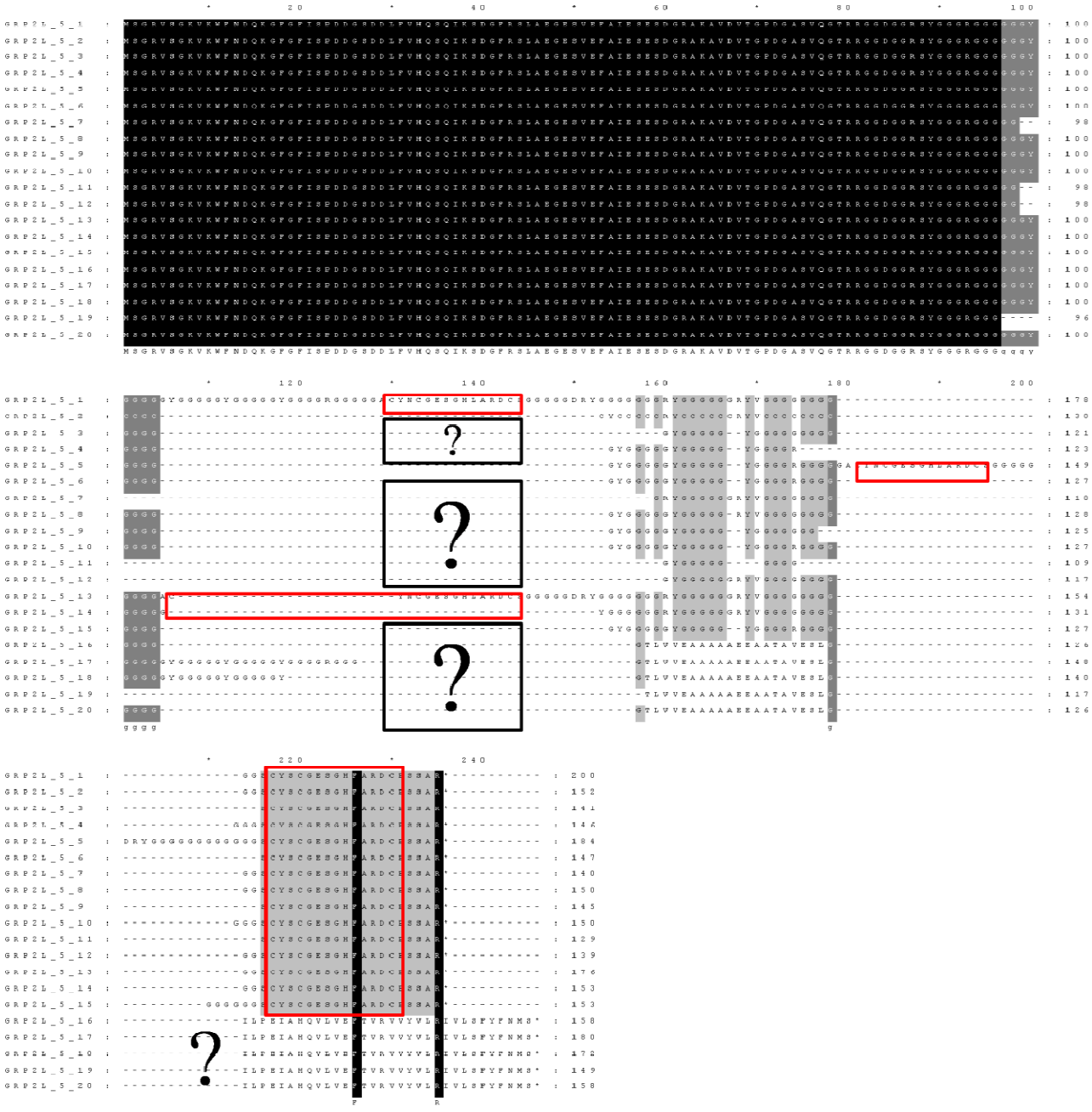

Fig. S4 Alignment of amino acids sequences predicted from alternative transcript of *GRP2L_5*. The black–gray scale background show the conservation of amino acids among sequences, the more darkness more conserved (conserved percent in black is 100%, in dark gray is 80% and in light gray is 60%). In the first row is present the CSD, red boxes indicates the presence of CCHC and the "?" indicates the lack of CCHC. The number of alternative transcripts accordingly to Phytozome is after the gene name *GRP2L_5*.

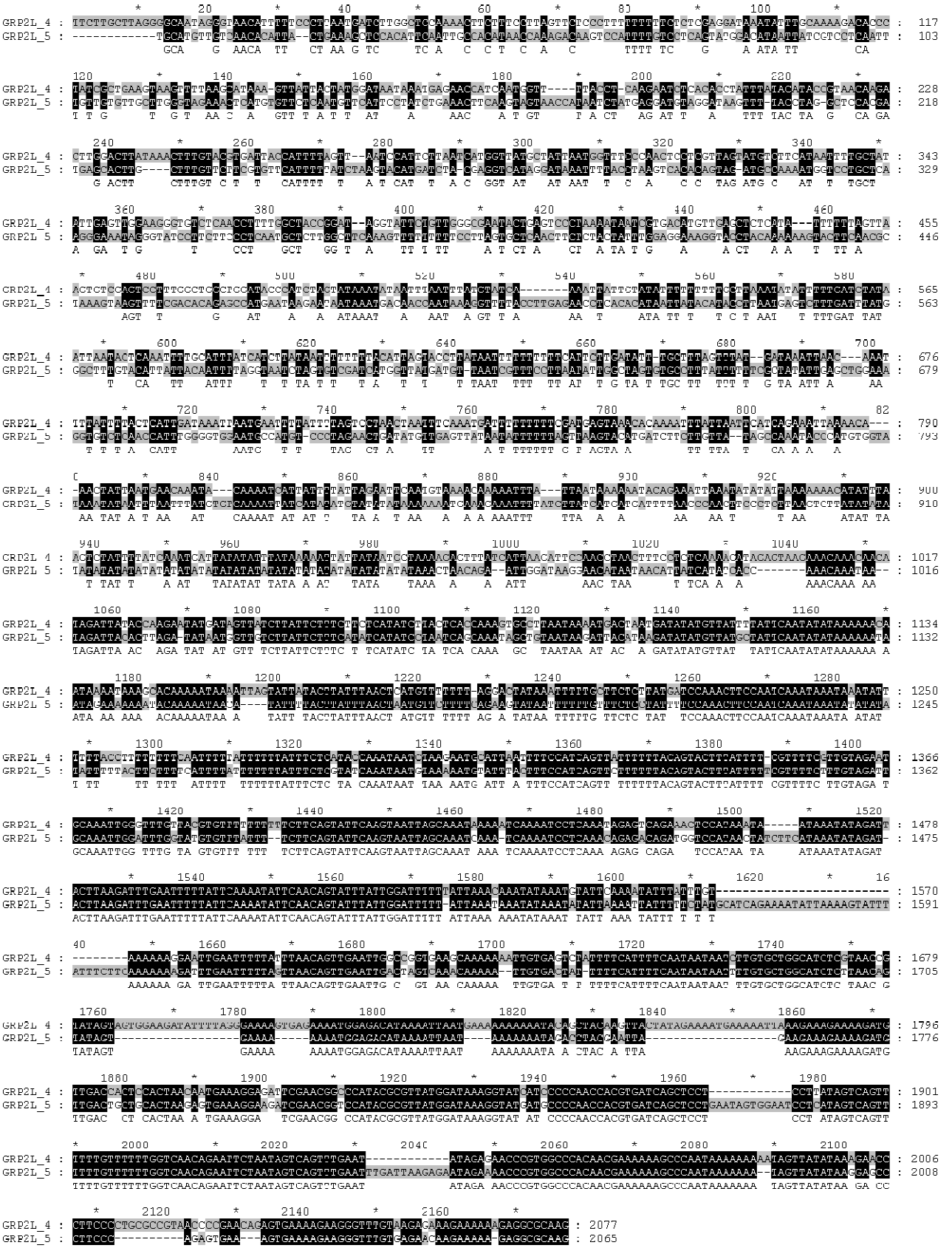
GRP2L_4
GRP2L_5

Fig. S5 Alignment of promoter regions of *GRP2L_4* and *GRP2L_5* (~2000 bp upstream the start codon). The initial translational codon (+1) is in the end of the sequences. The black background indicates identical nucleotides. The different nucleotides are shown by the gray and white backgrounds.

CAPÍTULO II

Soybean transformation for the expression of a glycine-rich, RNA-binding protein-encoding gene from *Arabidopsis thaliana* (*AtGRP2*)

Manuscrito em preparação

Soybean transformation for the expression of a glycine-rich, RNA-binding protein-encoding gene from *Arabidopsis thaliana* (*AtGRP2*)

Liane B. Poersch[1]; Beatriz Wiebke-Strohm[1;] Márcia Margis-Pinheiro[1] and Maria Helena Bodanese-Zanettini[1]✉

[1] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Brazil.

✉ maria.zanettini@ufrgs.br or mhbzanettini@yahoo.com.br, phone: 0055-51-33086725, fax: 0055-51-33089823, address: Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, CP 15053, 91501-970, Porto Alegre, RS, Brasil.

Abstract: Flowering is controlled by several environmental and developmental factors. Photoperiod and temperature are crucial for flowering in soybean and other plants. Tropical regions have a short photoperiod thus causing an early flowering and consequently a reduced productivity. An enormous set of GRP have been identified in various organisms and they can play several roles. In Arabidopsis, functional analysis of the *AtGRP2* gene (AT4g38680) showed that it is involved in seed and flower development, as well as at flowering time regulation. This gene can also play a role in cold acclimation. The study of genes that control the transition from vegetative to reproductive stage has great importance, being a requirement for eventual genetic manipulation of this trait. The goal of the present study was to produce transgenic soybean plants expressing the *AtGRP2* gene. Secondary somatic embryos of Bragg, IAS-5 and BRSMG 68 Vencedora soybean cultivars were used to introduce the gene into the soybean genome by two methods: (1) particle bombardment and (2) the combined DNA-free particle bombardment and *Agrobacterium* system. Six independent transformation events of Bragg cultivar were confirmed by PCR and thirty-one plants were regenerated from them. In the present moment the plants are under development in glass flasks. Two of these lines were originated from bombarded cluster and four lines resulted from bombardment/*Agrobacterium* system.

Key-words: soybean transformation; *AtGRP2*; glycine-rich proteins; cold-shock domain; RNA-binding proteins; flowering time.

Introduction

Soybean [*Glycine Max* (L.) Merril] is one of the most important cultures worldwide as a predominant source of both animal and human feed protein and cooking oil. Because of this remarkable status, soybean culture is the goal of a great number of breeding programs. However, the majority of these programs have been based on traditional breeding methods and therefore limited by the narrow genetic diversity that composes soybean germplasm (Priolli et al. 2002). This situation is worsened by the sexual incompatibility in interspecific and intergeneric crosses (Hu and Zanettini 1995; Bodanese-Zanettini et al. 1996). Genetic transformation provides the advantage to introduce new genes of interest, broadening the possible genetic combinations.

Flowering is controlled by several environmental and developmental factors. The complexity of this regulation is determined by an intricate network of signal transduction, through which the apical meristem is converted into a floral meristem (Zik and Irish 2003).Photoperiod and temperature are crucial for the occurrence of flowering in soybean, as they will cause qualitative changes in crop development (Major et al. 1975). Under constant photoperiod, the temperature has greater importance on flowering time, so that lower temperature increases the period for flowering (Embrapa 2003).

The soybean is originally a short-day plant. Most cultivars are better adapted to specific latitudes and sowing period due to their sensitivity to changes in photoperiod (Hartwig and Kiilhl 1979). Tropical regions have a short photoperiod thus causing an early flowering and consequently a reduced productivity (Shanmugasundaram and Tsou 1978).

An enormous set of **G**lycine-**R**ich **P**roteins (GRPs) have been identified in various organisms and they can play several roles, both structural and functional. A common characteristic of GRPs is the high content of glycine residues. These proteins are organized into four classes according to their primary structure and presence of additional domains.

Analyzing the *Arabidopsis thaliana*'s *AtGRP2* gene (AT4g38680), Fusaro et al. (2007) found that it is involved in seed and flower development, as well as at the flowering time regulation. When the expression of this gene was down-regulated in *A. thaliana*, the plant set

flowers precociously. The *AtGRP2* was also shown to be regulated by low temperatures and was expressed preferentially in meristematic and developing tissues under cell division. This gene encodes a nucleo-cytoplasmic GRP able to bind RNA that possesses three distinct domains: one **C**old-**S**hock **D**omain (CSD) containing the RNP1 consensus sequence in the N-terminal half, a **G**lycine-**R**ich region (GR) and two retroviral-type CCHC zinc finger (CCHC) interspersed in GR region. Due to these characteristics, the AtGRP2 protein was classified in class IV GRP (RNA-binding proteins). Additional studies with proteins containing CSD have shown that these proteins act like RNA chaperones and their expression are induced by low temperatures suggesting a role in cold acclimation (Jiang et al. 1997; Karlson et al. 2002).

The study of genes that control the transition from vegetative to reproductive stage has great importance being a requirement for eventual genetic manipulation of this trait. The goal of the present study is to produce transgenic soybean plants expressing the *AtGRP2* gene in order to evaluate its effect on the seed and flower development, as well as in the flowering period.

Material and Methods

Plant material and culture conditions

For transformations experiments Bragg, IAS-5 and BRSMG 68 Vencedora were used. Bragg and IAS-5 soybean cultivars were chosen due to their high susceptibility to *Agrobacterium tumefaciens* wild strains (Droste et al. 1994), as well as their response capacity to *in vitro* culture conditions (Droste et al. 2001; Droste et al. 2002; Droste et al. 2010). BRSMG 68 Vencedora has been shown to be also a genotype with high potential for somatic embryogenesis (Droste et al. 2010). Bragg and IAS-5 are North American-adapted cultivars, commonly used in genetic improvement programs, whereas BRSMG 68 Vencedora is a cultivar released by Brazilian breeding programs. BRSMG 68 Vencedora and IAS5 has commonly been indicated for commercial cropping in Brazil (MAPA 2009).

The use of secondary somatic embryos induced and proliferated *in vitro* as explants in experiments of transgenic soybean allow the regeneration of completely transformed plants due the unicellular origin of material (Sato et al. 1993). Embryogenic tissue from immature cotyledons was obtained according to the protocol described by Wiebke-Strohm et al. (2011). Briefly, the immature pods of growth field plants were superficially sterilized and the cotyledons placed in D40 medium, containing MS salts (Murashige and Skoog 1962), B5 vitamins (Gamborg et al. 1968), 40 mg/L **2,4-d**iclorofenoxiacetic (2-4D), 3% sucrose, 0,3% Phytagel® and pH 7.0 prior to autoclaving. The embryos were then transferred to D20 medium (D40 medium with 20 mg/L 2,4-D, pH 6.4) and sub-cultured every 14 days. The embryogenic cultures were maintained in 16/8 hours light/dark at light intensity of 22.5 $\mu Em^{-2}s^{-1}$.

Vector construction and bacteria culture conditions

The entry vector *pENTR Directional TOPO*® (Invitrogen) containing the *AtGRP2* gene was kindly provided by Dr. Gilberto Sachetto-Martins, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. This vector was introduced in *Escherichia coli* XL1 Blue strain (Stratagene) for multiplication and maintenance. The pENTR-*AtGRP2* was recombined with the super-expression destiny vector pH7WG2D,1 (Karimi et al. 2002) using the Gateway® system. This system relies on recombination of LR sites using the LR clonase™ II enzyme (Invitrogen). The recombination was performed accordingly manufacturer's instructions. The pH7WG2D,1-*AtGRP2* vector was introduced by electroporation in *E. coli* TOP 10 strain (Invitrogen) for multiplication and maintenance. The same procedure was used to transform the *A. tumefaciens* LBA4404 strain.

The T-DNA region of the final vector (pH7WG2D,1-*AtGRP2*) contains the hygromycin–phosphotransferase marker gene (*hpt*), the green fluorescent protein reporter gene (*gfp*) and the *AtGRP2* gene. The *AtGRP2* is driven by the cauliflower mosaic virus 35S (CAMV35S) promoter and terminator.

The bacteria culture conditions and the antibiotics used for plasmid/bacteria selection are shown in Table 1. The culture media used in this study were Luria Broth Base (Miller's LB Broth Base®) and LB, powder (Lennox L Agar), Invitrogen.

The **P**olimerase **C**hain **R**eaction (PCR) was used to confirm the insertion of the transgene into the plasmids, as well as the presence of plasmid in bacteria. The PCR primers used are provided in Table 2. The PCR amplification was carried out as follow: five minutes at 94 °C followed by 31 cycles of 94 °C for 45 seconds, 52 or 60 °C for 45 seconds, 72 °C for 45 seconds, with a final extension step of five minutes at 72 °C. The reaction mixture consisted of 2.5 µL PCR buffer 10X, 1.25 µL $MgCl_2$ 50 mM, 1 µL dNTP mixture 10 mM, 1 µL forward primer 10 mM, 1 µL reverse primer 10 mM, 0.2 µL *Taq* DNA Polimerase (5U/µL) (Invitrogen, São Paulo, Brazil) and 1 µL of DNA template, in a final volume of 25 µL. The positive constructions were sequenced to verify the transgenes integrity.

Transformation procedure

Two different methods were used to transform the soybean embryogenic tissue: (i) particle bombardment as described by Droste et al. (2002) and (ii) the combined DNA-free particle bombardment and *Agrobacterium* system as suggested by Wiebke-Strohm et al. (2011).

Briefly, in the transformation experiment by bombardment, 48 hours before the transformation 15 embryo clusters (about 0.67 mg/cluster) were placed in a Petri dish containing D20 medium. The Petri dishes were maintained uncovered in a laminar flow hood for 15 minutes, just before bombardment, to reduce the turgidity of the plant material (Vain et al. 1993). Bombardments were performed using a Particle Inflow Gun – PIG (Finer et al. 1992), with a pressure of 60 PSI and partial vacuum of 28 mmHg. Each Petri dish was bombarded once with M10 tungsten particles (Dupont, Wilmington, DE) coated with the pH7WG2D,1-*AtGRP2* plasmid (1 µg/µl). Ten dishes were used per cultivar.

In the bombardment/*Agrobacteirum* system DNA-free tungsten particles were used to induce wounds in the tissue prior to inoculation with *Agrobacterium*. Microwounds caused by tungsten particles enhance the *Agrobacterium* penetration and effective transformation (Bidney et al. 1992; May et al. 1995; Abdollahi et al. 2009). The bombardment procedure was the same as above described. Following bombardment, the embryo clusters were inoculated with *Agrobacterium* suspension harboring the pH7WG2D,1-*AtGRP2*, for 20 minutes. Inoculated explants were blotted on sterile filter paper and co-cultured for 48 hours on D20 medium

supplemented with 100 µM acetosyringone. Ten dishes with 15 embryo clusters each were used per cultivar.

Selection and regeneration of transgenic clones

The bombarded tissue remained in the same D20 medium for 12 days. After that the embryo clusters were transferred to D20 medium containing 12.5 mg/L of the selective agent hygromycin-B (Invitrogen) for 21 days. Subsequently, the embryos were cultured for three months on the same medium containing 25 mg/L hygromycin-B.

After the co-cultivation period, tissues submitted to the bombardment/*Agrobacterium* transformation procedure were washed in sterile distilled water, blotted on sterile filter paper and transferred to D20 medium containing 250 mg/L cefotaxime (Claforan®, Hoechst Marion Roussel) and 250 mg/L vancomycin (Vanclomin®, Teuto Brasileiro). Embryo clusters were maintained on this medium for 10 days and then were transferred to fresh D20 medium supplemented with the same antibiotics plus 12.5 mg/L hygromycin-B for more 21 days. Thereafter, plant material was kept on D20 medium containing the same three antibiotics, but with a concentration of 25 mg/L hygromycin-B. After 18 days, embryogenic tissues were transferred to fresh medium containing only hygromycin-B for additional 51 days, with subcultures every 14 days.

The following steps were the same for both transformation methods. After the selection period, pieces of green tissue were individually subculture every 14 days for two months in D20 medium without antibiotic in order to proliferate the hygromycin-resistant embryos. In the last 14 proliferation days 50 µM **Ab**scisic **A**cid (ABA) was added to the medium in order to increase the plant regeneration (Weber et al. 2007; Angoshtari et al. 2009; Droste et al. 2010).

To stimulate histodifferentiation the hygromycin-resistant embryo clusters were transferred to plates containing modified MSM6 maturation medium (Finer and McMullen 1991) composed of MS salts (Murashige and Skoog 1962), B5 vitamins (Gamborg et al. 1968), 6% sucrose, 1% activated charcoal, 50 µM ABA and 0.3% Phytagel® (pH 6.4 prior autoclaving). After 30 days,

histodifferentiated somatic embryos were individualized and transferred to the same medium without charcoal and ABA for one additional month.

The histodifferentiated embryos were placed on empty sterile dishes without medium for six hours in order to promote partial desiccation and thereby increase conversion velocity and frequency (Buchheim et al. 1989; Finer and McMullen 1991). To promote germination and conversion the embryos were transferred to MS0 conversion medium containing MS salts, B5 vitamins, 3% sucrose and 0,3% Phytagel®, pH 6.4 prior to autoclaving. Germination refers to root and shoot emission, while conversion was recorded as the development of the branched root and formation of at least one trifoliolate leaf (Walker and Parrot 2001). The regenerated plants were placed into 100 mL flasks containing 25 mL of the same medium.

DNA extraction

The histodifferentiated embryos that did not fully converted into plants were removed from culture medium and were immediately frozen in liquid nitrogen and stored at -80 °C. Histodifferentiated embryos derived from the same individual piece of hygromycin-resistant tissues were considered as an individual line and were used as a set for DNA extraction. The DNA extraction was performed according (Doyle and Doyle 1987) with modifications. DNA samples of each individual line were assayed for transgene presence. PCR conditions were performed as previously described.

Results

Vector construction

The confirmation of the insertion of the correct sequence of *AtGRP2* nucleotides in the destiny vector and its appropriate orientation was obtained by PCR and DNA sequencing. The

presence and integrity of pH7WG2D,1-*AtGRP2* plasmid extracted from *E. coli* were visualized in agarose gel electrophoresis under **u**ltra**v**iolet (UV) light (Fig 1a). Two PCRs were performed using: (i) the *AtGRP2* primer pair (Fig 1b) and (ii) the *hpt* primer pair (Fig 1c). The PCR positive pH7WG2D,1-*AtGRP2* was sequenced using the *AtGRP2* forward, *AtGRP2* reverse and P*35S* forward primers confirming the insertion of the transgene in the desirable orientation and also certifying nucleotide sequence integrity (data not shown).

The *A. tumefaciens* LBA4404 strain was electroporated to transform it with the pH7WG2D,1-*AtGRP2* vector. The presence of the plasmid in bacteria was also confirmed by PCR (Fig 2).

Proliferation and selection of embryo clusters and regeneration of transgenic plants

Soybean somatic embryos formed from immature cotyledons were proliferated on D20 medium for seven months (Fig 3a and 3b). Secondary somatic embryo clusters were transformed by bombardment and by the combined DNA-free particle bombardment and *Agrobacterium* system. After 90 days-selection, hygromycin-resistant green pieces were individually proliferated on D20 medium (Fig 3c). Each individual green piece of tissue detached from different sites at the transformed clusters was considered a putative independent transformation event. The individualized clusters were proliferated for two months on the same medium. Embryos/plants derived from an independent piece of hygromycin-resistant tissue (line) were considered clones. The total number of hygromycin-resistant clusters is shown in Table 3.

To stimulate histodifferentiation, hygromycin-resistant proliferated clusters were transferred to the MSM6 modified maturation medium (Fig 3d). After 30 days, somatic embryos were individualized and transferred to fresh maturation medium without charcoal and ABA for further 30 days. After that the individual histodifferentiated embryos were partially desiccated for six hours in empty Petri dishes (Fig 3e and 3f) and transferred to regeneration medium until the conversion into plants (Fig 3g).

The number of histodifferentiated embryos and recovered plants obtained from each transformation method are shown in Table 3. Converted plants were placed into glass flasks

containing the same medium (Fig 3h and 3i). Using bombardment method, 31 plants from 13 independent transformation events of the Bragg cultivar were regenerated. For cultivars IAS-5 and BRSMG 68 Vencedora, plants were not obtained by this method. Although a large number of histodifferentiated embryos of Bragg and IAS-5 cultivars resulted from transformation with the bombardment/*Agrobacterium* system, plants were not recovered. For BRSMG 68 Vencedora, the few hygromycin-resistant clusters not even formed histodifferentiated embryos.

Screening for transgenic events

PCR was used to screen histodifferentiated embryos that did not fully converted into plants representing all 62 lines. The P*35S* forward and *AtGRP2* reverse primer pair was used to confirm the integration of promoter and transgene into the soybean genome. As shown in Fig 4, no amplification was detected in untransformed embryos (negative control), whereas positive control (pH7WG1D,1-*AtGRP2* plasmid) and six lines presented the expected ~850 bp fragment. Two of these lines (Bomb Bragg 5.4.1 and Bomb Bragg 4.5.3) were originated from bombarded cluster and four lines (Agro Bragg 5.1.2, Agro Bragg 5.1.1, Agro Bragg 4.1 and Agro Bragg 5.2.3) resulted from bombardment/*Agrobacterium* system. The lines that did not produce PCR products were considered “escapes”.

Discussion

Proliferating somatic embryos have been utilized as target tissue for soybean transformation in numerous laboratories (Finer and McMullen 1991; Sato et al. 1993; Stewart et al. 1996; Schmidt et al. 2008). In our laboratory, many transgenic plants were obtained by bombardment using that target tissue (Droste et al. 2002; Weber 2007; Wiebke-Strohm 2010). Recently, Wiebke-Strohm et al. (2011) described, for the first time, the successful recovery of soybean transgenic fertile plants obtained from the combination of DNA-free particle

bombardment and *Agrobacterium* mediated transformation using clusters of somatic embryos as target.

It is long known that all soybean transformation methods described in the literature present problems such as low efficiency, poor reproducibility and limited cultivar specificity (Trick et al. 1997; Somers et al. 2003). In the present study, transgenic somatic embryos representing a total of six independent lines were successfully recovered from both transformation methods utilized. Recent reports (Homrich et al. 2008; Hernandez-Garcia et al. 2009; Kita et al. 2010; Wiebke-Strohm et al. 2011) using the same target tissue and transformation procedures have described an average number of transgenic events (from four to eleven) similar to the one obtained in this study.

The development of genetically engineered soybean has been limited to a small number of cultivars that respond to the *in vitro* stimulus (Tomlin et al. 2002). In previous studies with cultivars adapted to Brazilian environment conditions, IAS-5, Bragg and BRSMG 68 Vencedora were identified as suitable genotypes for successful embryogenesis and regeneration (Santos et al. 1997; Droste et al. 2001; Körbes and Droste 2005; Weber et al. 2007; Droste et al. 2010). Although a high number of histodifferentiated embryos was obtained from cultivars Bragg and IAS-5, plants were generated only from Bragg submitted to the bombardment transformation experiment. Surprisingly, in the present study, the IAS-5 cultivar has not regenerated plants, opposing results obtained by Droste et al. (2002), Weber et al. (2007) and Wiebke-Strohm et al. (2011) that have reported IAS-5 as the most responsive cultivar to regeneration. Although recent advances to improve embryo maturation and therefore its conversion into plants (Schmidt et al. 2005; Weber et al. 2007; Droste et al. 2010), soybean regeneration remains an "art" that requires considerable staff training in order to to develop the required condition to generate enough transgenic plants for writing a thesis or article for publication. Furthermore, the frequency of transgenic events that regenerate from each initial explant is often low (Somers et al. 2003).

In the *Agrobacterium* method, there is another obstacle besides the response of soybean to *in vitro* stimulus: the interaction between host and bacteria. There are few studies showing soybean susceptibility to *Agrobacterium*, with genotype-depending response (Delzer et al. 1990; Bailey et al. 1994; Droste et al. 1994). The compatible interaction between the Bragg, IAS-5, BR4 and CEP16 soybean cultivars and *A. tumefaciens* strains have been demonstrated by Droste et al.

(1994). Among these cultivars, Bragg was the most sensitive one to all the virulent *A. tumefaciens* strains tested, while IAS-5 was the least susceptible cultivar (Droste et al. 1994; Droste et al. 2010). This susceptible interaction is in agreement with our results, in which the stable integration of the T-DNA harboring the *AtGRP2* gene into the embryo genome was confirmed only for Bragg lines.

A high number of hygromycin-resistant embryos have been not confirmed for transgene stable integration and were considered as "escapes". The recovery of untransformed samples ("escapes") was also recorded for soybean transformation experiments in previous reports, although in a lower percentage (Olhoft et al. 2003; Homrich et al. 2008; Wiebke-Strohm et al. 2011). It is important to point out, that the hygromycin-based selection is the most trustful method to recover soybean transformed tissues and previous works showed that its application drastically reduced the number of escapes (reviewed by Somers et al. 2003).

What the present study still lacks is to obtain fertile mature transgenic plants over-expressing the gene *AtGRP2*. After that, phenotypes analyses, such as flower and seed development and timing, will provide information on the gene function in soybean and will make it possible to compare to that function described for Arabidopsis. Additional transformation experiments are currently in progress to allow regeneration of a higher number of transgenic lines/plants.

References

Abdollahi MR, Corral-Martínez P, Mousavi A, Salmanian AH, Moieni A, Seguí-Simarro JM (2009) An efficient method for transformation of pre-androgenic, isolated *Brassica napus* microspores involving microprojectile bombardment and *Agrobacterium*-mediated transformation. Acta Physiol Plant 31:1313-1317

Angoshtari R, Afshari RT, kalantari S, Omidi M (2009) Effects of abscisic acid on somatic embryogenesis and induction of desiccation tolerance in *Brassica Napus*. Asian Jl Plant Sci 8:276-284

Bailey MA, Boerma HR, Parrott WA (1994) Inheritance of *Agrobacterium tumefaciens*-induced tumorigenesis of soybean. Crop Sci 34:514-519

Bidney D, Scelonge C, Martich J, Burrus M, Sims L, Huffman G (1992) Microprojectile bombardment of plant tissues increases transformation frequency by *Agrobacterium tumefaciens*. Plant Mol Biol 18:301-313

Bodanese-Zanettini MH, Lauxen MS, Richter SNC, Cavalli-Molina S, Lange CE, Wang PJ, Hu CY (1996) Wide hybridization between Brazilian soybean cultivars and wild perennial relatives. Theor Appl Genet 93:703-709

Buchheim JA, Colburn SM, Ranch JP (1989) Maturation of soybean somatic embryos and the transition to plantlet growth. Plant Physiol 89:768-775

Delzer BW, Somers DA, Orf JH (1990) *Agrobacterium tumefaciens* susceptibility and plant regeneration of 10 soybean genotypes in maturity groups 00 to II. Crop Science 30:320-322

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15

Droste A, Bodanese-Zanettini MH, Mundstock E, Hu C-Y (1994) Susceptibility of Brazilian soybean cultivars to *Agrobacterium tumefaciens*. Brazil. J. Genetics 17:83-88

Droste A, Leite PCP, Pasquali G, Mundstock ECd, Bodanese-Zanettini MH (2001) Regeneration of soybean via embryogenic suspension culture. Scienti Agricola 58:753-758

Droste A, Pasquali G, Bodanese-Zanettini MH (2002) Transgenic fertile plants of soybean [*Glycine max* (L.) Merril] obtained from bombarded embryogenic tissue. Euphytica 127:367-376

Droste A, Silva AMd, Souza IFd, Wiebke-Strohm B, Bücker-Neto L, Bencke M, Sauner MV, Bodanese-Zanettini MH (2010) Screening of Brazilian soybean genotypes with high potential for somatic embryogenesis and plant regeneration. Pesq Agropec Bras 45:715-720

Embrapa - Empresa Brasileira de Pesquisa Agropecuária (2003) Tecnologias de produção de soja região central do Brasil. In: http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Soja/SojaCentralBrasil2003/exigencias.htm

Finer JJ, McMullen MD (1991) Transformation of soybean via particle bombardment of embryogenic suspension culture tissue. In Vitro Cell Dev Biol 27:175

Finer JJ, Vain P, Jones MW, McMullen MD (1992) Development of the particle inflow gun for DNA delivery to plant cells Plant Cell Reports 11:323-328

Fusaro AF, Bocca SN, Ramos RL, Barroco RM, Magioli C, Jorge VC, Coutinho TC, Rangel-Lima CM, De Rycke R, Inze D, Engler G, Sachetto-Martins G (2007) AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. Planta 225:1339-1351

Gamborg OL, Miller RA, Ojima K (1968) Nutrient requirements of suspension cultures of soybean root cells. Exp Cell Res 50:151-158

Hartwig EE, Kiilhl RAS (1979) Identification and utilization of a delayed flowering characterization in soybeans for short-day conditions. Field Crops Research 2:145-151

Hernandez-Garcia CM, Martinelli AP, Bouchard RA, Finer JJ (2009) A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Plant Cell Rep 28:837-849

Homrich MS, Passaglia LMP, Pereira JF, Bertagnolli PF, Pasquali G, Zaidi MA, Altosaar I, Bodanese-Zanettini MH (2008) Resistance to *Anticarsia gemmatalis* Hübner (Lepidoptera, Noctuidae) in transgenic soybean (*Glycine max* (L.) Merrill Fabales, Fabaceae) cultivar IAS5 expressing a modified Cry1Ac endotoxin. Gen Mol Biol 31:522-531

Hu CY, Zanettini MHB (1995) Embryo culture and embryo rescue from wide cross hybrids. In: Gamborg OL, Phillips GC (eds) Plant Cell, Tissue and Organ Cultures: Fundamental Methods. Springer, Berlin, pp 129-141

Jiang W, Hou Y, Inouye M (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. J Biol Chem 272:196-202

Karimi M, Inze D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci 7:193-195

Karlson D, Nakaminami K, Toyomasu T, Imai R (2002) A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. J Biol Chem 277:35248-35256

Kita Y, Nakamoto Y, Takahashi M, Kitamura K, Wakasa K, Ishimoto M (2010) Manipulation of amino acid composition in soybean seeds by the combination of deregulated tryptophan biosynthesis and storage protein deficiency. Plant Cell Rep 29:87-95

Körbes AP, Droste A (2005) Carbon sources and polyethylene glycol on soybean somatic embryo conversion. Pesq. Agropec. Bras. 40:211-216

Major DJ, Johnson DR, Tanner JW, Anderson IC (1975) Effects of daylenght and temperature on soybean development. Crop Science 15:174-179

MAPA - Ministério da Agricultura e Pecuária (2009) Zoneamento agrícola de risco climático: cultivares de soja ano-safra 2008/2009. In: www.agricultura.gov.br

May GD, Afza R, Mason HS, Wiecko A, Novak FJ, Arntzen CJ (1995) Generation of transgenic banana (*Musa acuminata*) plants via *Agrobacterium*-mediated transformation. Bio/Technol 13:486-492

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473-398

Olhoft PM, Flagel LE, Donovan CM, Somers DA (2003) Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method. Planta 216:723-735

Priolli RHG, Mendes-Junior CT, Arantes NE, Contel EPB (2002) Characterization of Brazilian soybean cultivars using microssatellite markers. Genet Mol Biol 25:185-193

Santos KGB, Mundstock E, Bodanese-Zanettini MH (1997) Genotype-specific normalization of soybean somatic embryogenesis through the use of an ethylene inhibitor. Plant Cell Rep 16:859-864

Sato S, Newell C, Kolacz K, Tredo L, Finer J, Hinchee M (1993) Stable transformation via particle bombardment in two different soybean regeneration systems. Plant Cell Rep 12:408-413

Schmidt MA, LaFayette PR, Artelt BA, Parrott WA (2008) A comparison of strategies for transformation with multiple genes via microprojectile-mediated bombardment. In Vitro Cell Dev Biol 44:162-168

Schmidt MA, Tucker DM, Cahoon EB, Parrott WA (2005) Towards normalization of soybean somatic embryo maturation. Plant Cell Rep 24:383-391

Shanmugasundaram S, Tsou SCS (1978) Photoperiod and critical duration for flower induction in soybean. Crop Science 18:598-601

Somers DA, Samac DA, Olhoft PM (2003) Recent advances in legume transformation. Plant Physiol 131:892-899

Stewart CN, Jr., Adang MJ, All JN, Boerma HR, Cardineau G, Tucker D, Parrott WA (1996) Genetic transformation, recovery, and characterization of fertile soybean transgenic for a synthetic *Bacillus thuringiensis* cryIAc gene. Plant Physiol 112:121-129

Tomlin ES, Branch SR, Chamberlain D, Gabe H, Wright MS, JR. CNS (2002) Screening of soybean, *Glycine max* (L.) Merril, lines for somatic embryo induction and maturation capacbility from immature cotyledons. In Vitro Cell. Dev. Biol.-Plant 38:543-548

Trick HN, Dinkins RD, Santarém ER, Di R, Samoylov V, Meurer CA, Walker DR, Parrot WA, Finner JJ, Collins GB (1997) Recent advances in soybean transformation. Plant Tiss Cult Biotechnol 3:9-26

Vain P, McMullen MD, Finer JJ (1993) Osmotic treatment enhances particle bombardment-mediated transient and stable transformation of maize. Plant Cell Reports 12:84-88

Walker Dr, Parrot WA (2001) Effect of polyethylene glycol and sugar alcohols on soybean somatic embryo and conversion. Plant Cell Tissue and Organ Culture 64:55-62

Weber RLM (2007) Transformação genética de soja [*Glycine max* (L) Merril] com um gene que codifica uma osmotina de *Solanum nigrum* var. *americanum*, visando a resistência de moléstias fungícas. Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências. Universidade Federal do Rio Grande do Sul, Porto Alegre, 67 p

Weber RLM, Körbes AP, Baldasso DA, Callegari-Jacques SM, Bodanese-Zanettini MH, Droste A (2007) Beneficial effect of abscisic acid on soybean somatic embryo maturation and conversion into plants. Plant Cell Cult Micropropag 3:1-9

Wiebke-Strohm B (2010) Obtenção de plantas estavelmente transformadas pelo sistema integrado bombardeamento/*Agrobacterium* e análise funcional dos genes que codificam as ureases estruturais da soja. Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética. Universidade Federal do Rio Grande do Sul, Porto Alegre, 129 p

Wiebke-Strohm B, Droste A, Pasquali G, Osorio MB, Bücker-Neto L, Passaglia LMP, Bencke M, Homrich MS, Margis-Pinheiro M, Bodanese-Zanettini MH (2011) Transgenic fertile soybean plants derived from somatic embryos transformed via the combined DNA-free particle bombardment and *Agrobacterium* system. Euphytica 177:343-354

Zik M, Irish VF (2003) Flower Development: Initiation, Differentiation, and Diversification. Annu Rev Cell Dev Biol 19:119-140

Table 1 Bacteria culture conditions and antibiotic used for plasmid/bacteria selection.

Bacteria	Plasmid	[ ] mg/L – antibiotic	Culture time	Growth temperature
E. coli XL1	pENTR-*AtGRP2*	50 – kan	12-17 h	37 °C
E. coli TOP 10	pH7WG2D,1-*AtGRP2*	20 – Sm 75 – Sp	12-17 h	37 °C
A. Tumefaciens LBA4404	pH7WG2D,1-*AtGRP2*	20 – Sm 75 – Sp 25 – Rif	48 h	28 °C

Kan: kanamycin; Sm: Streptomycin; Sp: Spectinomycin; Rif: Rifampycin

Tabela 2 Primer set used to detect the transgene and plasmid presence

Target	Orientation	Nucleotide sequence	PCR product size (pb)	Melting temperature
AtGRP2	Forward	5'ATGAGCGGAGACAACGGCGGTG3'	612	60 °C
	Reverse	5'TTAACGTCCACCGCTGGTG3'		
hpt	Forward	5'GGCCTCCAGAAGAAGATGTTG3'	453	52 °C
	Reverse	5'GAGCCTGACCTATTGCATCTCC 3'		
P*35S*	Forward	5'CGCACAATCCCACTATCCTT3'	139*	52 °C

* Fragment length was considered from primer until the end of the promoter sequence.

Table 3 Soybean hygromycin-resistant clusters, histodifferentiated embryos and converted plants obtained from transformation experiments

Cultivar	Method					
	Bombardment			Bombardment/*Agrobacterium*		
	Bragg	**IAS-5**	**BRSMG 68 Vencedora**	**Bragg**	**IAS-5**	**BRSMG 68 Vencedora**
Number of clusters submitted to transformation	150	150	150	150	150	150
Selection						
Number of hygromycin-resistant clusters	58	17	ND	29	34	6
Histodifferentiation						
Number of lines	35	2	0	15	10	0
Number of embryos	1782	3	0	455	562	0
Conversion						
Number of lines	13	0	0	0	0	0
Number of plants	31	0	0	0	0	0

ND: Non determined.

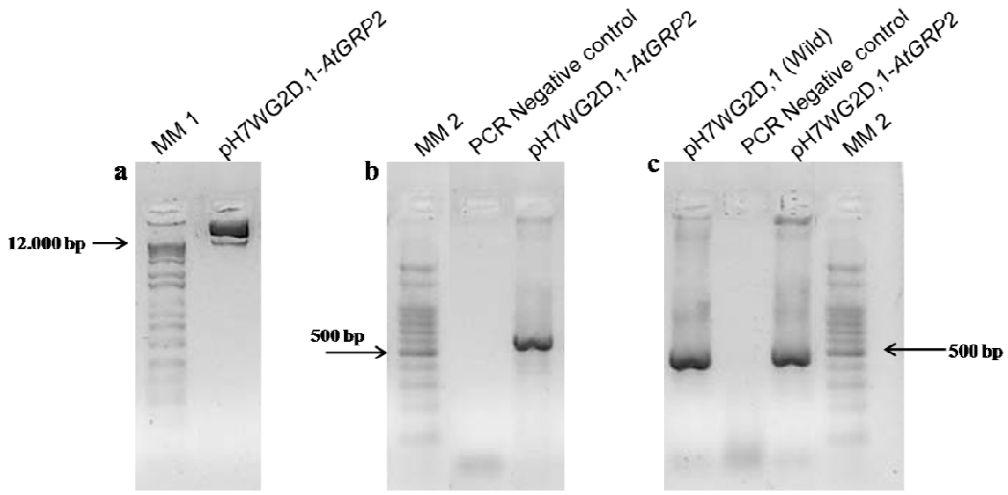

Fig. 1 Confirmation of *AtGRP2* integration into the destiny vector pH7WG2D,1 by PCR. (a) Purified pH7WG2D,1-*AtGRP2* plasmid from *E. coli* TOP 10 strain. (b) PCR carried out with purified pH7WG2D,1-*AtGRP2* plasmid using *AtGRP2* primer pair, which amplifies the *AtGRP2* encoding region. (c) PCR carried out with purified pH7WG2D,1-*AtGRP2* plasmid using *hpt* primer pair, which amplifies the *hpt* gene present in the vector. MM1: Molecular Marker, Ladder 1 Kb Plus Invitrogen; MM2: Molecular Marker: Ladder 100 bp Ludwig. 1% agarose gels stained with ethidium bromide and visualized under UV light.

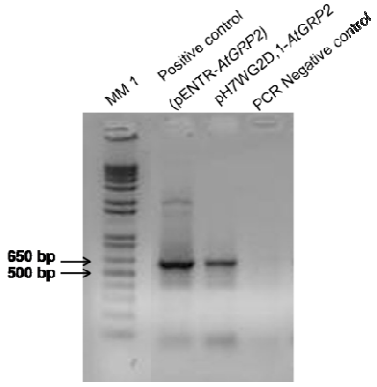

Fig. 2 Confirmation of presence of the pH7WG2D,1-*AtGRP2* destiny vector in *A. tumefaciens* by PCR using *AtGRP2* primer pair. MM1: Molecular Marker, Ladder 1 Kb Plus Invitrogen. 1% agarose gel stained with ethidium bromide and visualized under UV light.

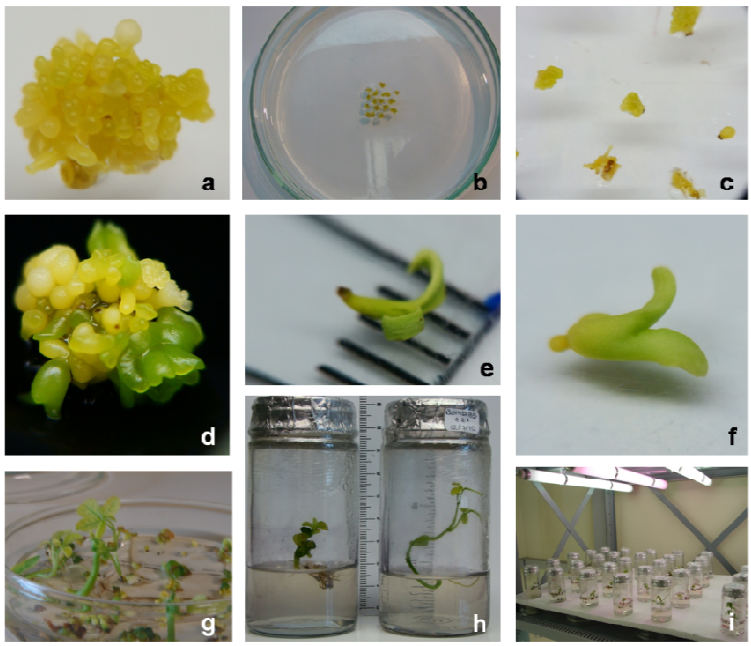

Fig. 3 Soybean somatic embryos on proliferation, selection, histodifferentiation, and conversion. (a) Secondary somatic embryo cluster (two mm in diameter and around 0,67 mg) used as target for transformation process. (b) Petri dish with 15 embryo clusters prepared for transformation. (c) Hygromycin-resistant clusters in D20 proliferation medium. (d) Initial embryo histodifferentiation one week after transference to modified MSM6 maturation medium. (e-f) Desiccation of histodifferentiated embryos. (g) Germinated embryos on MS0 regeneration medium. (h-i) Converted plants in glass flasks with MS0 medium.

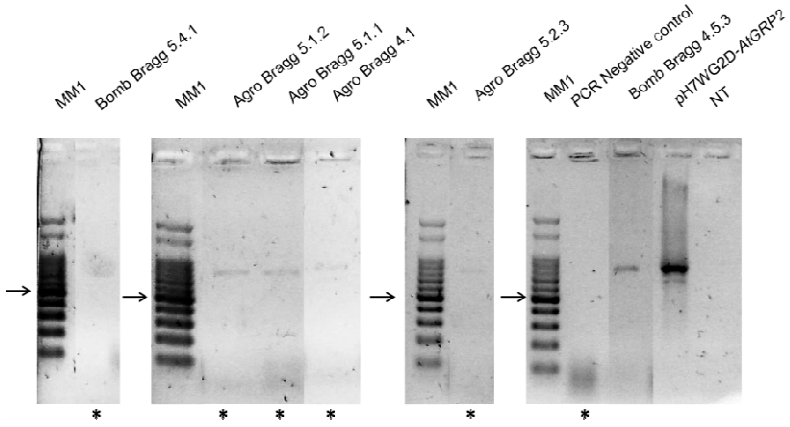

Fig. 4 Detection of promoter and transgene in soybean embryos. PCR products amplified from DNA templates extracted from histodifferentiated embryos not fully converted into plants: embryos from six transgenic lines (*); pH7WG2D,1-*AtGRP2* plasmid (positive control); NT (untransformed embryos). Bomb (Bombardment transformation method); Agro (Bombardment/*Agrobacterium* transformation system); Numbers after cultivar names represent the different events (lines); MM1: Molecular Marker, Ladder 100 bp, Ludwing; arrows indicates 500 bp band. 1% agarose gel, stained with ethidium bromide visualized under UV light.

DISCUSSÃO GERAL

5 DISCUSSÃO GERAL

Recentemente, a soja (cultivar Williams 82) teve o sequenciamento de seu genoma praticamente concluído, representando 950 do total estimado de 1.115 **m**ega**b**ases (Mb). A poliploidia teve um profundo efeito na estruturação do genoma da soja. Estudos indicaram que houve dois eventos de duplicação total do genoma, seguidos pela diversificação e perda gênica, bem como rearranjos entre cromossomos, resultando em um genoma com aproximadamente 75% dos genes presentes em múltiplas cópias (Schmutz et al. 2010). As ferramentas de sequenciamento *high-throughput* e a bioinformática vêm permitindo uma análise global de expressão gênica, mudando o foco de um único gene para todo o genoma. Consequentemente, uma análise mais ampla onde todos os genes de uma família ou que codifiquem proteínas relacionadas pode ser realizada. Por meio destas ferramentas, foi possível a identificação de 47 genes que codificam proteínas ricas em glicina ligantes de RNA no genoma da soja.

O presente trabalho, além da identificação das GRPs ligante de RNA de soja, traz como novidade três supostos motivos presentes em quatro proteínas da subclasse IVa (Glyma08g44150, Glyma08g44170, Glyma18g08590 e Glyma18g08610) sem função descrita na literatura. Também de forma pioneira foi realizada a análise filogenética da classe IV de GRPs presente no genoma desta leguminosa. Com este resultado pode-se observar que este grupo de proteínas ainda está em curso de evolução e possivelmente possui papéis redundantes no genoma paleopoliplóide da soja. Ainda, pode-se verificar que a classificação das proteínas baseada apenas na presença de domínios não mostra a história evolutiva destes genes. Para uma melhor compreensão do papel e ação destas proteínas na planta deve-se levar em conta a classificação atual baseada na estrutura juntamente com a divisão filogenética.

As proteínas ligantes de RNA podem exercer suas funções combinando diversos domínios de ligação ao RNA, frequentemente auxiliados por domínios adicionais (os quais podem aumentar a "afinidade" pelo RNA ou promover interações proteína-proteína). Estas proteínas podem empacotar o RNA, protegê-lo, organizá-lo e prepará-lo para processos pós-transcricionais. Além disso, são capazes de coordenar vários passos na expressão gênica pelo fato de sua multifuncionalidade (Elliott and Ladomary 2010).

Inúmeros estudos têm demonstrado que os domínios presentes na classe IV de GRPs são extremamente conservados mesmo entre organismos filogeneticamente distantes (Burd and Dreyfuss 1994; Albà and Pagès 1998; Graumann and Marahiel 1998; Maruyama et al. 1999; Karlson et al. 2002; Karlson and Imai 2003; Nakaminami et al. 2006; Chaikam and Karlson 2010). Provavelmente existiu uma proteína

contendo um CSD ancestral no começo da evolução de organismos unicelulares, 3,5 bilhões de anos atrás. A presença de CSD em bactérias e eucariotos reafirma esta idéia (Graumann and Marahiel 1996). Em procariotos as proteínas contendo RRM tendem a conter apenas um destes domínios, enquanto proteínas de eucariotos frequentemente contem múltiplos RRMs que estão associados a domínios adicionais. Sua importância na adaptação a estresses tem sido mantida praticamente inalterada por milhões de anos (Elliott and Ladomary 2010).

Resultados prévios mostraram que o gene *GRP2* de *Arabidopsis thaliana* (*AtGRP2*), está envolvido em processos de desenvolvimento de sementes, na estrutura da flor, bem como no tempo para que ocorra a floração, podendo ainda ter um papel na aclimatação ao frio (Fusaro et al. 2007). Com base nestes resultados prévios, no presente trabalho uma maior atenção foi dispensada à subclasse IVc, que inclui as proteínas com um CSD, pelo seu suposto papel na aclimatação ao frio.

Proteínas que contém um CSD são bem conservadas e possuem um papel crítico para a adaptação de bactérias a baixas temperaturas. A função destas proteínas, nesta condição, em procariotos tem sido extensivamente estudada e está bem estabelecida (Jones et al. 1987; Goldstein et al. 1990; Jiang et al. 1997; Yamanaka et al. 1998; Wang et al. 1999; Xia et al. 2001; Phadtare and Severinov 2005; Nakaminami et al. 2006). Entretanto, esta relação é pobremente entendida em eucariotos, principalmente em plantas. De um modo geral, proteínas com CSD em eucariotos superiores desenvolvem diversas funções (Graumann and Marahiel 1998; Chaikam and Karlson 2008).

Proteínas de plantas contendo o CSD são frequentemente relacionadas com a aclimatação ao frio uma vez que algumas são fortemente induzidas em resposta a baixas temperaturas (de Oliveira et al. 1990; Nakaminami et al. 2006; Fusaro et al. 2007; Kim et al. 2007; Chaikam and Karlson 2008). Entretanto, o papel funcional e a importância na resposta a este estresse permanecem desconhecidos (Chaikam and Karlson 2008). Até o momento, nenhuma evidência experimental demonstrando um papel na aclimatação *in planta* foi descrita.

Além de seu papel na adaptação ao frio, proteínas de plantas contendo o CSD também estão implicadas em diversos aspectos do desenvolvimento de diferentes sistemas modelo. No mutante sensível ao frio (cepa BX04 de *E. coli*) foram observados defeitos na divisão celular e em baixas temperaturas as células se tornaram alongadas (Xia et al. 2001). Em camundongos, um *knock-out* do gene codificante da proteína YB-1 provocou letalidade do embrião (Lu et al. 2005; Uchiumi et al. 2006). Outro polipeptideo, Lin 28, de *Caenorhabditis elegans* controla transições dos primeiros estágios desenvolvimento (Moss et al. 1997).

Conforme os resultados obtidos neste trabalho com os experimentos de expressão tanto *in silico* quanto *in vivo* pode-se inferir que as proteínas de soja contendo CSD estão de alguma maneira envolvidas no desenvolvimento deste organismo, uma vez que a maioria dos genes desta família possui uma expressão aumentada em tecidos em desenvolvimento e sob divisão celular. Apesar de conterem um CSD, a expressão de dois dos seis genes codificantes de proteínas da subclasse IVc é reprimida quando a planta é exposta a baixas temperaturas. Esta observação sugere um envolvimento destes genes em outros processos. Outro indício de que proteínas que contém o CSD estejam envolvidas em diversos níveis de controle de expressão gênica é fornecido pelo fato de que o este domínio possui uma estrutura tridimensional parecida ao do RRM e OB (*oligonucleotide/oligossacaride fold*) e também sendo capaz de se ligar a ácidos nucléicos (Elliott and Ladomary 2010).

Com a liberação de sequências de genomas de diferentes plantas, promotores vegetais estão recebendo aumentada atenção como um dos primeiros reguladores da expressão gênica. Análises realizadas com os promotores dos seis genes da subclasse IVc identificados em soja reforçam os resultados obtidos com os dados de expressão em diferentes situações.

Após todas as informações discutidas neste trabalho um fato interessante constitui a identidade das proteínas codificadas pelos genes *GRP2L_4* e *GRP2L_5* da subclasse IVc, além deles compartilharem uma região promotora muito parecida. Tendo em vista que estes genes encontram-se em cromossomos diferentes e parecem apresentar o mesmo padrão de expressão, fica a pergunta se esta foi uma duplicação recente ou se estas regiões do DNA foram mantidas quase inalteradas por uma pressão de seleção específica. Isto poderá ser respondido com as metodologias atuais de seqüenciamento, onde seria interessante (não só para este par específico de genes mas para todos os membros da classe) seqüenciar outras variedades de soja e fazer uma comparação dos genes entre as cultivares. A longo prazo, seria importante também a comparação de sequências de espécies diferentes para uma melhor compreensão do comportamento dos genes ao longo da evolução vegetal.

Plantas transgênicas têm sido usadas extensivamente para estudos de expressão e função gênica. Entretanto na maioria destes estudos plantas modelo tais como Arabidopsis, tabaco e arroz são utilizadas. No presente trabalho cultivares de soja, utilizadas em programas de melhoramento e recomendadas para plantio no Brasil, foram transformadas com o gene *AtGRP2* para estudo funcional do mesmo. No vetor de super-expressão utilizado, este gene está sobre controle constitutivo do promotor CAMV35S. O gene *AtGRP2* foi escolhido por dois motivos: (i) seus efeitos fenotípicos de super-expressão e silenciamento já foram descritos em Arabidopsis e (ii) por pertencer a subclasse IVc das GRPs ligantes de RNA e ter sido usado como isca representante para a busca de ortólogos em soja. Além disto, na época da realização do experimento de transformação genética os genes endógenos da soja para esta subclasse ainda não haviam

sido identificados. No momento, as plantas encontram-se em etapa de regeneração e assim que plantas transgênicas maduras forem obtidas, estas serão submetidas a análises de expressão e fenotípica, tanto na geração T0 quanto na T1.

Estudos genéticos sobre o tempo de floração, na planta modelo *A. thaliana*, mostram mais de uma via regulando este processo (Mouradov et al. 2002; Simpson and Dean 2002). Fusaro et al. (2007) mostram que mudanças na expressão de *AtGRP2* influenciam o tempo de floração, a estrutura da flor, bem com o desenvolvimento da semente neste organismo. Contudo, ainda não está claro como o produto deste gene está agindo. Os autores levantam duas hipóteses: (i) esta proteína estaria controlando o tempo de floração em um nível pós-transcricional ou (ii) a proteína poderia estar agindo a nível traducional como uma chaperona de RNA, assim como WCSP1 e CSPA.

Desta forma será interessante investigar o efeito da super-expressão do gene *AtGRP2* em soja e verificar se o mesmo afeta o processo de floração. Como mencionado na introdução deste trabalho, o fotoperíodo e a temperatura possuem grande influência no tempo de floração em soja. Geralmente baixas temperaturas promovem um florescimento mais tardio, enquanto altas temperaturas promovem um florescimento mais precoce. Uma variação muito grande na temperatura e em estágios não apropriados durante o desenvolvimento da soja causa enormes prejuízos para a produção (http://www.cnpso.embrapa.br/index.php?op_page=38& cod_pai=1&cod_noticia=338). Registros de florescimento precoce nas lavouras de soja têm sido feitos no Brasil nos últimos anos devido a aumento na temperatura (Embrapa 2009; Embrapa 2010a; Embrapa 2010b).

Tendo em vista o aquecimento global, variações de temperaturas anuais e influências climáticas cada vez mais diversificadas como La niña e El niño, o estudo de genes que podem interferir no tempo de floração torna-se importante, principalmente genes endógenos. A disponibilidade de uma soja contendo um gene que influencia o tempo de floração sem influência da temperatura seria interessante.

A identificação de genes codificantes de GRPs ligantes de RNA em soja foi realizada. A partir deste ponto, pesquisas devem ser realizadas com cada subclasse identificada. A respeito da subclasse IVc uma caracterização inicial sobre a estrutura e perfil de expressão em diferentes órgãos da planta e em diversos estresses foi realizada neste trabalho. Entretanto, este foi apenas um passo inicial para o entendimento do papel destas proteínas em soja. Futuros trabalhos verificando como estas proteínas se comportam no mutante de *E. coli* sensível ao frio, bem como experimentos para testar a atividade chaperona de RNA destas proteínas devem ser realizados. Pesquisas adicionais para confirmar a localização subcelular predita destas proteínas também são importantes, bem como verificar a presença *in vivo* de cada transcrito alternativo dos

genes *GRP2L_4* e *GRP2L_5*. Outro ponto que não deve ser esquecido é uma caracterização inicial sobre o perfil de expressão do gene *GRP2L_1*, que não foi realizada no presente trabalho.

Quanto aos experimentos de transgenia, após (ou em paralelo) a caracterização fenotípica das plantas da geração T0 e T1 contendo o gene *AtGRP2,* experimentos de transformação genética de soja com os genes endógenos desta subclasse deveriam ser realizados. Baseado em dados da literatura sobre funções de proteínas contendo CSD e nos perfis de expressão apresentados neste trabalho, dentre os seis genes identificados seria interessante começar os estudos utilizando os genes *GRP2L_3* e *GRP2L_4/5*, os quais parecem estar agindo no desenvolvimento e em resposta ao frio e ataque de fungos de maneira oposta.

REFERÊNCIAS DA INTRODUÇAO E DISCUSSÃO GERAL

Albà MM, Pagès M (1998) Plant proteins containing the RNA-recognition motif. Trends in Plant Science 3:15-21

Bailey MA, Boerma HR, Parrot WA (1993) Genotype effetcs on proliferative embryogenesis and plant regeneration of soybean. In Vitro Cell Dev Biol 29P:102-108

Barros R (2009) Doenças da cultura da soja. Tecnologia e Produção: Soja e Milho 2008/2009. In: www.fundacaoms.org.br/request.php?97, pp 109-122

Bocca SN, Magioli C, Mangeon A, Junqueira RM, Cardeal V, Margis R, Sachetto-Martins G (2005) Survey of glycine-rich proteins (GRPs) in the *Eucalyptus* expressed sequence tag database (ForEST). Genetics and Molecular Biology 28:608-624

Bodanese-Zanettini MH, Lauxen MS, Richter SNC, Cavalli-Molina S, Lange CE, Wang PJ, Hu CY (1996) Wide hybridization between Brazilian soybean cultivars and wild perennial relatives. Theor Appl Genet 93:703-709

Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science 265:615-621

Chaikam V, Karlson D (2008) Functional characterization of two cold shock domain proteins from *Oryza sativa*. Plant Cell Environ 31:995-1006

Chaikam V, Karlson DT (2010) Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. BMB Rep 43:1-8

CNPSO - Centro Nacional de Pesquisa de Soja (2009a) In: http://www.cnpso.embrapa.br/index.php?op_page=22&cod_pai=16 Acess: 2009

CNPSO - Centro Nacional de Pesquisa de Soja (2009b) In: http://www.cnpso.embrapa.br/index.php?op_page=294&cod_pai=17 Acess: 2009

CONAB – Companhia Nacional de abastecimento (2010) In: www.conab.gov.br/conteudos.php?a=1252&t=2 Acess: 2009

de Oliveira DE, Seurinck J, Inze D, Van Montagu M, Botterman J (1990) Differential expression of five Arabidopsis genes encoding glycine-rich proteins. Plant Cell 2:427-436

Dita MA, Rispail N, Prats E, Rubiale D, Singh KB (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. Euphytica 147: 1-24

Droste A, Pasquali G, Bodanese-Zanettini MH (2000) Integrated Bombardment and *Agrobacterium* Transformation System: an Alternative Method for Soybean Transformation. Plant Mol Biol Rep 18:51-59

Droste A, Pasquali G, Bodanese-Zanettini MH (2002) Transgenic fertile plants of soybean [*Glycine max* (L.) Merril] obtained from bombarded embryogenic tissue. Euphytica 127:367-376

Elliott D, Ladomary M (2010) Molecular Biology of RNA. Oxford University Press, USA

Embrapa – Empresa Brasileira de Pesquisa Agropecuária (2003) Tecnologias de produção de soja região central do Brasil. In: http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Soja/SojaCentralBrasil2003/exigencias.htm

Embrapa – Empresa Brasileira de Pesquisa Agropecuária (2008) Sistemas de produção 13 - Tecnologias de produção de soja - Região Central do Brasil 2009 e 2010. Londrina, 261 p

Embrapa – Empresa Brasileira de Pesquisa Agropecuária (2009) http://www.cnpso.embrapa.br/index.php?op_page=38&cod_pai=1&cod_noticia =338, acesso em: dezembro de 2009.

Embrapa – Empresa Brasileira de Pesquisa Agropecuária (2010a) http://www.embrapa.br/imprensa/noticias/2010/janeiro/1a-semana/embrapa-soja-esclarece-sobre-florescimento-precoce-da-soja-no-pr-sp-e-ms/, acesso em: janeiro de 2010.

Embrapa – Empresa Brasileira de Pesquisa Agropecuária (2010b) http://www.cnpso.embrapa.br/alerta/ver_alerta.php?cod_pagina_sa=214&cultura=1, acesso em: janeiro de 2010.

Finer JJ (1988) Apical proliferation of embryogenic tissue of soybean [*Glycine max* (L.) Merrill]. Plant Cell Rep 7:238-241

Finer JJ, McMullen MD (1991) Transformation of soybean via particle bombardment of embryogenic suspension culture tissue. In Vitro Cell Dev Biol 27:175

Finer JJ, Vain P, Jones MW, McMullen MD (1992) Development of the particle inflow gun for DNA delivery to plant cells. Plant Cell Reports 11:323-328

Fusaro A, Mangeon A, Junqueira RM, Rocha CAB, Coutinho TC, Margis R, Sachetto-Martins G (2001) Classification, expression pattern and comparative analysis of sugarcane expressed sequences tags (ESTs) encoding glycine-rich proteins (GRPs). Genet Mol Biol 24:263-273

Fusaro AF, Bocca SN, Ramos RL, Barroco RM, Magioli C, Jorge VC, Coutinho TC, Rangel-Lima CM, De Rycke R, Inze D, Engler G, Sachetto-Martins G (2007) AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. Planta 225:1339-1351

Fusaro AF, Sachetto-Martins G (2007) Blooming time for plant glycine-rich proteins. Plant Signal Behav 2:386-387

Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67:16-37

Gendra E, Moreno A, Alba MM, Pages M (2004) Interaction of the plant glycine-rich RNA-binding protein MA16 with a novel nucleolar DEAD box RNA helicase protein from Zea mays. Plant J 38:875-886

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the Cross-Legume Advances Through Genomics Conference. Plant Physiol 137:1228-1235

Goellner K, Loehrer M, Langenbach C, Conrath U, Koch E, Schaffrath U (2010) *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. Mol Plant Pathol 11:169-177

Goldstein J, Pollitt NS, Inouye M (1990) Major cold shock protein of Escherichia coli. Proc Natl Acad Sci U S A 87:283-287

Graumann P, Marahiel MA (1996) A case of convergent evolution of nucleic acid binding modules. Bioessays 18:309-315

Graumann PL, Marahiel MA (1998) A superfamily of proteins that contain the cold-shock domain. Trends Biochem Sci 23:286-290

Gutterson N, Zhang JZ (2004) Genomics applications to biotech traits: a revolution in progress? Curr Opin Plant Biol 7:226-230

Hanano S, Sugita M, Sugiura M (1996) Isolation of a novel RNA-binding protein and its association with a large ribonucleoprotein particle present in the nucleoplasm of tobacco cells. Plant Mol Biol 31:57-68

Hartwig EE, Kiihl RAS (1979) Identification and utilization of a delayed flowering characterization in soybeans for short-day conditions. Field Crops Research 2:145-151

Hiraga S, Minakawa H, Takahashi K, Takahashi R, Hajika M, Ohtsubo KHN (2007) Evaluation of somatic embryogenesis from immature cotyledons of Japanese soybean cultivars. Plant Biotecnol 24:435-440

Hofmann N, Nelson RL, Korban S (2004) Influence of media components and pH on somatic embryo induction in three genotypes of soybean. Plant Cell, Tissue and Organ Culture 77:157-163

Homrich MS, Passaglia LMP, Pereira JF, Bertagnolli PF, Pasquali G, Zaidi MA, Altosaar I, Bodanese-Zanettini MH (2008a) Resistance to *Anticarsia gemmatalis* Hübner (Lepidoptera, Noctuidae) in transgenic soybean (*Glycine max* (L.) Merrill Fabales, Fabaceae) cultivar IAS5 expressing a modified Cry1Ac endotoxin. Gen Mol Biol 31:522-531

Homrich MS, Passaglia LMP, Pereira JF, Bertagnolli PF, Salvadori JR, Nicolau M, Kaltchuk-Santos E, Alves LB, Bodanese-Zanettini MH (2008b) Agronomic performance, chromosomal stability and resistance to velvetbean caterpillar of transgenic soybean expressing *cry1Ac* gene. Pesq Agropec Bras 43:801-807

Horsch RB, Fry JE, Hoffman NL, Eichholtz D, Rogers SG, Fraley RT (1985) A simple and general method for transfering genes into plants. Science 227:1229-1231

Hu CY, Zanettini MHB (1995) Embryo culture and embryo rescue from wide cross hybrids. In: Gamborg OL, Phillips GC (eds) Plant Cell, Tissue and Organ Cultures: Fundamental Methods. Springer, Berlin, pp 129-141

Jiang W, Hou Y, Inouye M (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. J Biol Chem 272:196-202

Jones PG, VanBogelen RA, Neidhardt FC (1987) Induction of proteins in response to low temperature in *Escherichia coli*. J Bacteriol 169:2092-2095

Karlson D, Imai R (2003) Conservation of the cold shock domain protein family in plants. Plant Physiol 131:12-15

Karlson D, Nakaminami K, Toyomasu T, Imai R (2002) A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. J Biol Chem 277:35248-35256

Kim JY, Park SJ, Jang B, Jung CH, Ahn SJ, Goh CH, Cho K, Han O, Kang H (2007) Functional characterization of a glycine-rich RNA-binding protein 2 in *Arabidopsis thaliana* under abiotic stress conditions. Plant J 50:439-451

Kim YO, Kim JS, Kang H (2005) Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. Plant J 42:890-900

Kingsley PD, Palis J (1994) GRP2 proteins contain both CCHC zinc fingers and a cold shock domain. Plant Cell 6:1522-1523

Kohli A, Twyman RM, Abranches R, Wegel E, Stoger E, Christou P (2003) Transgene integration, organization and interaction in plants. Plant Mol Biol 52:247-258

Kwak KJ, Kim YO, Kang H (2005) Characterization of transgenic Arabidopsis plants overexpressing GR-RBP4 under high salinity, dehydration, or cold stress. J Exp Bot 56:3007-3016

Lu ZH, Books JT, Ley TJ (2005) YB-1 is important for late-stage embryonic development, optimal cellular stress responses, and the prevention of premature senescence. Mol Cell Biol 25:4625-4637

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys 444:139-158

Major DJ, Johnson DR, Tanner JW, Anderson IC (1975) Effects of daylenght and temperature on soybean development. Crop Science 15:174-179

Mangeon A, Junqueira RM, Sachetto-Martins G (2010) Functional diversity of the plant glycine-rich proteins superfamily. Plant Signal Behav 5:99-104

Maruyama K, Sato N, Ohta N (1999) Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. Nucleic Acids Res 27:2029-2036

Moss EG, Lee RC, Ambros V (1997) The cold shock domain protein Lin-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. Cell 88:637-646

Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. Plant Cell 14 Suppl:S111-130

Nakaminami K, Karlson DT, Imai R (2006) Functional conservation of cold shock domains in bacteria and higher plants. Proc Natl Acad Sci U S A 103:10122-10127

Phadtare S, Severinov K (2005) Nucleic acid melting by *Escherichia coli* CspE. Nucleic Acids Res 33:5583-5590

Pierce BA (2004) GENETICA: Um enfoque conceitual. Guanabara Koogan S.A., Rio de Janeiro

Priolli RHG, Mendes-Junior CT, Arantes NE, Contel EPB (2002) Characterization of Brazilian soybean cultivars using microssatellite markers. Genet Mol Biol 25:185-193

Reetz ER, Snatos C, Corrêa S, Silveira D, Carvalho C, Beling RR (2008) Anuário Brasileiro da Soja. In: Beling RR (ed) Anuário Brasileiro da Soja 2008, 136 p

Rodrigues O, Didonet AD, Lhamby JCB, Luz PFBJSd (2001) Resposta quantitativa do florescimento da soja à temperatura e ao fotoperíodo. Pesq Agropec Bras 36:431-437

Sachetto-Martins G, Franco LO, de Oliveira DE (2000) Plant glycine-rich proteins: a family or just proteins with a common motif? Biochim Biophys Acta 1492:1-14

Sato S, Newell C, Kolacz K, Tredo L, Finer J, Hinchee M (1993) Stable transformation via particle bombardment in two different soybean regeneration systems. Plant Cell Rep 12:408-413

Schmidt MA, LaFayette PR, Artelt BA, Parrott WA (2008) A comparison of strategies for transformation with multiple genes via microprojectile-mediated bombardment. In Vitro Cell Dev Biol 44:162-168

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 XC, 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 palaeopolyploid soybean. Nature 463:178-183

Shanmugasundaram S, Tsou SCS (1978) Photoperiod and critical duration for flower induction in soybean. Crop Science 18:598-601

Simpson GG, Dean C (2002) Arabidopsis, the Rosetta stone of flowering time? Science 296:285-289

Somers DA, Samac DA, Olhoft PM (2003) Recent advances in legume transformation. Plant Physiol 131:892-899

Stanford JC (1988) The biobalist process. Trends in Biotech 6:299-302

Steinert PM, Mack JW, Korge BP, Gan SQ, Haynes SR, Steven AC (1991) Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. Int J Biol Macromol 13:130-139

Tollefson J (2010) Food: The global farm. Nature 466:554-556

Trick HN, Dinkins RD, Santarém ER, Di R, Samoylov V, Meurer CA, Walker DR, Parrot WA, Finner JJ, Collins GB (1997) Recent advances in soybean transformation. Plant Tiss Cult Biotechnol 3:9-26

Uchiumi T, Fotovati A, Sasaguri T, Shibahara K, Shimada T, Fukuda T, Nakamura T, Izumi H, Tsuzuki T, Kuwano M, Kohno K (2006) YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation. J Biol Chem 281:40440-40449

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:202-210

Wang N, Yamanaka K, Inouye M (1999) CspI, the ninth member of the CspA family of *Escherichia coli*, is induced upon cold shock. J Bacteriol 181:1603-1609

Weber RLM (2007) Transformação genética de soja [*Glycine max* (L) Merril] com um gene que codifica uma osmotina de *Solanum nigrum* var. *americanum*, visando a resistência de moléstias fungícas. Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências. Universidade Federal do Rio Grande do Sul, Porto Alegre, 67 p

Wiebke-Strohm B (2010) Obtenção de plantas estavelmente transformadas pelo sistema integrado bombardeamento/*Agrobacterium* e análise funcional dos genes que codificam as ureases estruturais da soja. Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética. Universidade Federal do Rio Grande do Sul, Porto Alegre, 129 p

Wiebke-Strohm B, Droste A, Pasquali G, Osorio MB, Bücker-Neto L, Passaglia LMP, Bencke M, Homrich MS, Margis-Pinheiro M, Bodanese-Zanettini MH (2011) Transgenic fertile soybean plants derived from somatic embryos transformed via the combined DNA-free particle bombardment and *Agrobacterium* system. Euphytica 177:343-354

Wiebke B, Ferreira F, Pasquali G, Bodanese-Zanettini MH, Droste A (2006) Influence of antibiotics on embryogenic tissue and *Agrobacterium tumefaciens* suppression in soybean genetic transformation. Bragantia 65:543-551

Xia B, Ke H, Inouye M (2001) Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in *Escherichia coli*. Mol Microbiol 40:179-188

Yamanaka K, Fang L, Inouye M (1998) The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. Mol Microbiol 27:247-255

Yang C, Zhao T, Yu D, Gai J (2009) Somatic embryogenesis and plant regeneration in Chinese soybean (*Glycine max* (L.) Merr.) — impacts of mannitol, abscisic acid, and explant age. In Vitro Cell Dev Biol 45:180-188

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* (L.) Merr.) genetic map and their association with EST markers. . Theor Appl Genet 108:1131-1139

Zik M, Irish VF (2003) Flower Development: Initiation, Differentiation, and Diversification. Annu Rev Cell Dev Biol 19:119-140