Lauro Bücker Neto

ESTUDO DO FATOR DE TRANSCRIÇÃO ASR5 EM PLANTAS DE ARROZ (*Oryza sativa*) E IDENTIFICAÇÃO DE PROTEÍNAS EM RESPOSTA AO ESTRESSE POR ALUMÍNIO EM *Arabidopsis thaliana*

Tese apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito para a obtenção do título de doutor em Genética e Biologia Molecular

Orientadora: Prof. Dra. Maria Helena Bodanese Zanettini

Coorientadora: Prof. Dra. Márcia Margis

Linha de Pesquisa: Mapeamento, identificação de genes, cultura de tecidos e transformação genética de plantas de interesse agronômico

Porto Alegre

2014

Lauro Bücker Neto

ESTUDO DO FATOR DE TRANSCRIÇÃO ASR5 EM PLANTAS DE ARROZ (*Oryza sativa*) E IDENTIFICAÇÃO DE PROTEÍNAS EM RESPOSTA AO ESTRESSE POR ALUMÍNIO EM *Arabidopsis thaliana*

Tese apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito para a obtenção do título de doutor em Genética e Biologia Molecular

BANCA EXAMINADORA

_____

Francismar Correa Marcelino Guimarães – EMBRAPA Soja

_____

Andreia Carina Turchetto-Zolet – Universidade Federal do Rio Grande do Sul

_____

Fernanda Stanisçuaski – Universidade Federal do Rio Grande do Sul

O trabalho aqui apresentado foi desenvolvido no laboratório de Genética Vegetal do Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS - Porto Alegre), em colaboração com o Prof. Dr. Zhiyong Wang do Carnegie Institution for Science – Department of Plant Biology (Stanford University – California, EUA).

Fonte financiadora

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES);

AGRADECIMENTOS

Agradeço à:
Professora Dra. Maria Helena Bodanese Zanettini (orientadora);
Professora Dra. Márcia Margis (coorientadora);
Professor Dr. Zhiyong Wang (orientador exterior);
Professora Dra. Luciane Maria Pereira Passaglia (coorientadora mestrado, coautora artigo);
Professora Dra. Andreia Carina Turchetto-Zolet (coautora artigo);
Professor Dr. Alexandro Cagliari (suporte experimental);
Professor Dr. Júlio Cesar de Lima (coautor artigo);
Professor Dr. Rogerio Margis (coautor artigo);
Dra. Beatriz Wiebke-Strohm (coautora artigo);
Dra. Graciela Castilhos (coautora artigo);
Dra. Shouling Xu (coautora artigo - espectrometria de massa);
Dra. Tingting Xiang (suporte experimental);
Dr. Chan Ho Park (suporte experimental);
Dr. Jim Guo (suporte físico);
Dr. Luiz Felipe Valter de Oliveira (coautor artigo);
Dr. Rafael Augusto Arenhart (coautor artigo);
Dr. Ricardo Luís Mayer Weber (coautor artigo);
Dr. Thomas Hartwig (coautor artigo - suporte experimental);
Msc. Bi Yang (suporte experimental e físico);
Msc. Caroline Cabreira (coautora artigo);
Msc. Chuangqi Wei (coautor artigo - espectrometria de massa);
Msc. Marina Borges Osorio (coautora artigo);
Msc. Marta Bencke (coautora artigo);
Msc. Rafael Rodrigues de Oliveira (coautor artigo);
Msc. Ronei Dorneles Machado (suporte experimental);
Msc. Shuolei Bu (suporte experimental);
Dra. Sunita Patil (suporte experimental);
Dasha Salvage (intervenções linguísticas);
Elmo Cardoso (tramites burocráticos);

RESUMO

As plantas são organismos sésseis que continuamente enfrentam situações ambientais adversas, o que acarreta em reduções significativas da biomassa e da produtividade. O trabalho, aqui exposto, teve como objetivo avaliar o papel dos fatores de transcrição ASR (do ingles *ABA*, *stress and ripening*) na resposta a estresses abióticos em plantas de arroz. Também teve como objetivo avaliar as respostas de plantas de *Arabidopsis thaliana* ao estresse produzido nos momentos iniciais da exposição ao metal alumínio. O capítulo 1 da presente tese, compara a expressão de miRNAs entre plantas silenciadas para o gene ASR5 (ASR5_RNAi) e plantas não transformadas (controle). De um total de 279 miRNAs maduros identificados, distribuídos em 60 famílias, 159 foram diferencialmente expressos quando as duas bibliotecas foram comparadas. Uma correlação negativa entre o MIR167 e seu gene alvo (LOC_Os07g29820) também foi confirmada por PCR em tempo real. Este é o primeiro trabalho sugerindo o envolvimento das proteínas ASR na regulação da expressão de miRNAs em planta. O segundo capítulo apresenta o estudo das proteínas ASR na manutenção da homeostase do pH em plantas de arroz. Verificou-se uma diminuição do crescimento radicular em plantas silenciadas em solução ácida, quando comparadas com plantas não transformadas nas mesmas condições. Também foi analisada a viabilidade da ponta de raízes quanto ao dano causado pelo baixo pH e diferentes concentrações de Ca^{+2}, demonstrando que a adição de $CaCl_2$ é capaz de aliviar o efeito tóxico do excesso de protons H^+. Diversos genes reprimidos nas plantas silenciadas e envolvidos no mecanismo de manutenção do pH em células vegetais, também foram investigados. O terceiro e último capítulo é dedicado ao estudo da resposta inicial de plantas de *Arabidopsis thaliana* ao estresse por alumínio. Plantas com 7 dias de idade foram expostas a uma concentração de 25 µM de $AlCl_3$ durante 3 horas e modificações na abundância de proteínas foi investigada com a técnica de espectrometria de massa. Um total de 3.213 proteínas foram identificadas, sendo que destas, 293 apresentaram variação no nível de expressão. Diversas proteínas com expressão induzida são funcionalmente associadas com a detoxificação de espécies reativas de oxigênio (ROS), indicando que o tratamento ocasionou estresse oxidativo nas raízes de *A. thaliana*. Também foram identificadas uma proteína mitocondrial carreadora de substrato e uma acyl-CoA oxidase com possível papel nos mecanismos de defesa em resposta a alumínio e com potencial para futuros estudos funcionais na planta modelo. De uma maneira geral, os resultados aqui apresentados mostram, pela primeira vez, que ASR5 está envolvida na regulação de miRNAs e na homeostase do pH em plantas de arroz, além de identificar proteínas responsivas ao estresse por alumínio em *A. thaliana*.

Palavras-chave: Proteínas ASR. Alumínio. *Oryza sativa*. *Arabidopsis thaliana*. miRNA

ABSTRACT

Plants are sessile organisms that continuously face adverse environmental situations, leading to a significant reduction in biomass and yield. The aim of the present work was to further study the ASR (ABA, stress and ripening) transcription factors in rice plants. Moreover, the responses of *Arabidopsis thaliana* to aluminum stress were also analyzed. The chapter 1 of this thesis compares the expression of mature miRNAs in the ASR5 silenced plants (ASR5_RNAi) and in non-transformed plants (control). From a total of 279 mature miRNA of 60 families, 159 were differentially expressed. A negative correlation of MIR167 and its target gene (LOC_Os07g29820) was also confirmed by real time RT-qPCR. This is the first report showing the involvement of ASR proteins in miRNA gene expression regulation. The second chapter presents the study of participation of ASR proteins in the maintenance of pH homeostasis in rice plants. The evaluation of root growth in ASR5_RNAi plants upon acid solution showed inhibition of root growth when compared to non-transformed plants in the same condition. Root tip feasibility and damage caused by low pH and different concentrations of Ca^{+2} was also analyzed. The results indicate that addition of $CaCl_2$ is capable of alleviating the toxic effects of H^+ protons. Several genes downregulated in silenced plants and involved in pH maintenance in plant cells have also been investigated. This work demonstrates the importance of ASR transcription factors in a biological process not yet described. The third and final chapter describes the study of the initial response of *Arabidopsis thaliana* to aluminum stress. Seven-day old seedlings were treated with 25 µM $AlCl_3$ for 3 hours and submitted to quantitative analyses by mass spectrometry. A total of 3,213 proteins were identified, from which 293 proteins were differentially responsive upon aluminum treatment. Several proteins with increased expression in response to the treatment are functionally associated with reactive oxygen species (ROS), indicating that the Al^{3+} exposure caused oxidative stress in the roots of *A. thaliana*. A mitochondrial substrate carrier (At1g78180) and an acyl-CoA oxidase (At3g51840) with a putative role in Al defense were also up-regulated and constitute interesting targets for functional studies of aluminum toxicity in the model plant. Overall, the results here presented show for the first time that ASR5 is involved in miRNA and pH homeostases regulation in rice plants and also identify proteins responsive to aluminum stress in *A. thaliana*.

Keywords: ASR proteins. Aluminum. *Oryza sativa*. *Arabidopsis thaliana*. miRNA

LISTA DE ABREVIATURAS

ABA - ácido abscísico

Al - Alumínio

cDNA - DNA complementar

Cv - cultivar

DNA - ácido desoxiribonucleico (do Inglês, *deoxiribonucleic acid*)

GA - giberilina (do Inglês, *gibberellin*)

µM - micromolar

mM - milimolar

PCR - reação em cadeia da DNA polimerase (do ingles, *polymerase chain reaction*)

PUGNAc - O-(2-acetamido-2-deoxy-D-glucopyranosylideneamino)*N*-phenylcarbamate

RNAi - RNA de interferência

RNAseq - sequenciamento de RNA (do ingles, *RNA sequencing*)

ROS - Espécies reativas de oxigênio (do ingles, *reactive oxigen species*)

RT-qPCR - Reação em cadeia da DNA polimerase quantitative precedida de transcrição reversa (do ingles, *reverse transcription quantitative* PCR)

s - segundos

Ssp - subespécie

SUMÁRIO

1 INTRODUÇÃO

Além do aumento da população mundial, existem diversas preocupações acerca do futuro da produção agrícola. A disponibilidade de terras aráveis está decrescendo em virtude de técnicas de manejo não sustentáveis, que, por sua vez, têm intensificado problemas como a erosão e a degradação do solo (STOCKING, 2003). Estudos recentes indicam que as mudanças climáticas globais afetarão seriamente o crescimento das mais variadas culturas de interesse agronômico, bem como a própria conservação das terras cultivadas (CHRISTENSEN et al., 2007; MEEHL et al., 2007). Ainda, Van Velthuizen et al. (2007) estimaram que somente 3,5% da área terrestre pode ser considerada totalmente livre de fatores limitantes ao crescimento vegetal.

Uma vez que existem limitações físicas, morfológicas e moleculares inerentes à habilidade de resposta das plantas, a superação dessas restrições passa pelo desenvolvimento e aplicação de novas tecnologias que visem, principalmente, o melhoramento das culturas em resposta aos mais variados estímulos ambientais. As modernas abordagens de estudos transcritômicos, metabolômicos e proteômicos, conjuntamente com análises integradas desses dados têm propiciado um melhor entendimento dos sistemas biológicos como um todo (CRAMER et al., 2011), mas a compreensão dos complexos mecanismos subjacentes ainda está distante de ser plenamente revelada.

1.1 TOXIDEZ POR ALUMÍNIO

Apesar de abundante na crosta terrestre (KOCHIAN et al., 2002), o alumínio encontra-se geralmente quelado a outros ligantes ou em formas não fitotóxicas como aluminosilicatos ou precipitados (DRISCOLL; SCHECHER, 1990). Entretanto, em solos com baixo pH (<5), a solubilidade do alumínio é intensificada e o metal torna-se um agente xenobiótico extremamente pernicioso e, consequentemente, fator limitante da produção agrícola. Estima-se que cerca de 50% dos solos aráveis do mundo são considerados ácidos (VON UEXKÜLL; MUTERT, 1995), um processo que ocorre naturalmente devido a exposição à chuva ácida ou à remoção de cátions

básicos do solo, mas que pode ser intensificado com o emprego de técnicas agrícolas inapropriadas (DELHAIZE; MA; RYAN, 2012). No Brasil, os solos chamados latossolos e argissolos ocupam aproximadamente 58% da área territorial e são caracterizados como profundos, altamente intemperizados, ácidos, de baixa fertilidade natural sendo, algumas vezes, saturados por alumínio (EMBRAPA, 2006).

Uma vez em solos ácidos, o alumínio passa a ser incorporado pelas plantas, interagindo com diferentes alvos tanto no apoplasto quanto no simplasto e interferindo nos mais variados processos celulares (MARON et al., 2008). A toxidez do metal passa a ser perceptível quando da inibição do crescimento da raiz que, consequentemente, prejudica a absorção de água e nutrientes (BARCELO; POSCHENRIEDER, 2002; FAMOSO et al., 2010), aumentando a sensibilidade da planta a estresses de outra natureza. Estudos indicam que a inibição do crescimento radicular decorre do dano ao DNA e consequente bloqueio celular, culminando na diferenciação do centro de quiescência (ROUNDS; LARSEN, 2008). Dessa forma, a sobrevivência das plantas em meio contendo altas concentrações de alumínio depende da existência de mecanismos de detoxificação externos (ou de resistência) e/ou internos (ou de tolerância) (MA et al., 2002). O primeiro caso inclui modificações da parede celular, permeabilização seletiva da membrana plasmática, aumento do pH da rizosfera, bem como exudação de ácidos orgânicos (AO) e compostos fenólicos (MARON et al., 2008). Malato, citrato e oxalato formam complexos no citosol ou na interface raiz-solo, protegendo o tecido radicular (MA; RYAN; DELHAIZE, 2001). Em Arabidopsis, 70% da resistência ao alumínio é condicionada pela atividade do malato secretado pelas raízes das plantas expostas ao metal (LIU et al., 2009). No segundo caso, a quelação do metal no citosol e a compartimentalização no vacúolo já foram descritas para algumas espécies (GREVENSTUK; ROMANO, 2013; JIAN ZHENG S; FENG MA J; MATSUMOTO, 1998; MA et al., 1997). Em uma minuciosa revisão, Magalhães (MAGALHAES, 2006) postula que os genes de tolerância a alumínio são conservados entre monocotiledôneas e dicotiledôneas. Com base nesse modelo, Arabidopsis e arroz consagram-se como excelentes ferramentas para o estudo de mecanismos de resistência e tolerância ao alumínio em plantas, uma vez que possuem genomas completamente sequenciados e recursos genéticos, tais como populações mutantes, disponíveis publicamente.

Sensitive to proton rhizotoxicity 1 (STOP1) em Arabidopsis e *Al^{3+}resistance transcription factor 1* (ART1) em arroz, são fatores de transcrição ortólogos (OHYAMA et al., 2013), identificados por análise de mutantes e caracterizados como componentes moleculares chave na expressão de genes em raízes submetidas a elevadas concentrações de alumínio. Em *A. thaliana* STOP1 foi inicialmente identificado em plântulas sensíveis ao baixo pH e, posteriormente, foi demonstrado ser fundamental na resposta da planta ao alumínio (IUCHI et al., 2007). Embora sua expressão não seja induzida pelo metal, ele é o regulador de, pelo menos, três importantes genes na resposta da planta a Al^{3+}. ALMT1 e MATE1 são proteínas envolvidas no efluxo de malato e citrato, respectivamente, responsáveis pela detoxificação externa de alumínio (LIU et al., 2009). ALS3 é um *half-type* transportador ABC regulado por STOP1 e está possivelmente envolvido no direcionamento de Al^{3+} para tecidos menos sensíveis ao metal (LARSEN et al., 2005). Apesar dos genes regulados por STOP1 contribuírem de maneira significativa na resistência a alumínio em Arabidopsis, pelo menos dois outros genes atuam independentemente desse fator de transcrição. ALS1 codifica uma proteína *membrane-spanning domain* de um transportador ABC localizado no tonoplasto (LARSEN et al., 2007), enquanto STAR1 codifica um domínio de ligação a ATP de um transportador ABC localizado na membrana plasmática. Embora nenhum desses genes seja induzido por alumínio e seu mecanismo de funcionamento permaneça desconhecido, mutantes com perda de função são sensíveis ao metal (HUANG; YAMAJI; MA, 2010).

Uma característica peculiar das plantas de arroz consiste na sua capacidade de tolerar concentrações elevadas de alumínio, quando comparadas a outros cereais (FAMOSO et al., 2010). Muito embora o mecanismo dessa resposta ainda não tenha sido esclarecido, genes chave têm sido identificados. *Al^{3+}resistance transcription factor 1* (ART1), um fator de transcrição do tipo dedo de zinco C2H2, foi caracterizado como fundamental na regulação da expressão de genes envolvidos na detoxificação do alumínio (YAMAJI et al., 2009). Seis genes regulados por ART1 já foram descritos. OsFRDL4 (*Ferric Reductase Defective Like* 4) é um transportador de citrato do tipo MATE (*multidrug and toxic compound extrusion*) responsável por parte da variação na tolerância entre diferentes genótipos de arroz (YOKOSHO; YAMAJI; MA, 2011). STAR1 codifica um domínio de ligação a nucleotídeo de um transportador ABC (*bacterial-type*), que interage com o domínio transmembrana de

um transportador ABC codificado por *STAR2*. Diferentemente do gene *STAR1* de Arabidopsis, a expressão do complexo composto pelas proteínas STAR1 e STAR2 (não identificado na planta modelo) em arroz é induzida em resposta ao alumínio, muito embora plantas mutantes também apresentem fenótipo de sensibilidade ao metal. Postula-se que estejam envolvidos no transporte de UDP-glucose para o apoplasto, onde o substrato atuaria modificando a parede celular e prevenindo o acúmulo de alumínio (HUANG et al., 2009). A proteína Nrat1 está envolvida com o transporte específico de alumínio para o meio intracelular (XIA et al., 2010). OsALS1 de arroz e AtALS1 de Arabidopsis são proteínas localizadas no tonoplasto, porém, o gene *OsALS1* é induzido em resposta ao alumínio e é expresso em todo o tecido radicular, enquanto *AtALS1* é constitutivamente expresso na tecido vascular, hidatódios e ápice da raiz (HUANG et al., 2012; LARSEN et al., 2007). Mais recentemente, Xia et al. (XIA; YAMAJI; MA, 2013) caracterizaram OsCDT3 como um pequeno peptídeo ancorado na membrana plasmática, cujo papel seria barrar a entrada de alumínio no simplasto, ligando-se diretamente ao metal e evitando os malefícios de sua toxicidade. O gene codificante da proteína OsCDT3 é expresso principalmente em raízes e induzido por Al^{+3}, mas não por pH ou outros metais. Plantas com nocaute do gene apresentaram menor tolerância ao alumínio, bem como um aumento na concentração do metal em vacúolos de células da raiz.

Em uma abordagem diferente, Arenhart et al. (ARENHART et al., 2013a) demonstraram que os níveis de expressão do gene *ASR5* (do inglês *absiscic acid*, *stress and ripening*) aumentam em resposta a alumínio e que plantas ASR5-RNAi foram incapazes de crescer em meio contendo o metal. Recentemente, foi provado que a proteína ASR5 também atua como regulador direto da expressão de STAR1 (ARENHART et al., 2014) e que, como ART1, também participa na regulação de genes de resposta ao alumínio.

1.2 *ARABIDOPSIS THALIANA*: EUDICOTILEDÔNEA MODELO DE ESTUDO VEGETAL

Arabidopsis thaliana é uma planta herbácea da família *Brassicaceae* largamente utilizada como organismo modelo para estudos de plantas nas áreas de pesquisa básica em genética, biologia celular e molecular. Apesar de não apresentar importância agronômica, possui relação filogenética com espécies cultivadas tais como o repolho (*Brassica oleraceae*, grupo Capitata) e o rabanete (*Raphanus*

sativus). O emprego de plantas de Arabidopsis nas mais variadas áreas de pesquisa acadêmica e aplicada decorre de uma série de características muito peculiares ao organismo. Esta espécie possui um genoma pequeno, de aproximadamente 125 Mpb, sequenciado e anotado (ARABIDOPSIS INITIATIVE, 2000), bem como mapas genéticos e físicos de todos os cromossomos (http://www.arabidopsis.org/servlets/mapper). O ciclo de vida é de aproximadamente 6 semanas, desde o período de germinação até a maturação das sementes, e o processo de polinização é eminentemente autogâmico. Cada planta é capaz de produzir cerca de 5000 sementes em um espaço restrito e com técnicas simples de cultivo (tanto *in vitro* quanto *ex vitro*). Por fim, eficientes protocolos de transformação utilizando *Agrobacterium tumefaciens*, bem como um amplo número de linhagens mutantes e a disponibilidade de tais informações (http://www.arabidopsis.org/index.jsp), fazem deste organismo um modelo para o estudo das plantas com flores.

1.3 ARROZ: MONOCOTILEDÔNEA MODELO DE ESTUDO

O arroz (*Oryza sativa*) é considerado um alimento de fundamental relevância na dieta de 2,4 bilhões de pessoas, atingindo uma produção mundial anual de 590 milhões de toneladas (EMBRAPA, 2014). No Brasil, a produção anual é estimada em 11,7 milhões de toneladas, sendo o Estado do Rio Grande do Sul o principal produtor nacional (IBGE, 2014). Além de sua inquestionável importância econômica, o arroz é considerado planta modelo de estudo para as monocotiledôneas, uma vez que possui o menor genoma entre os cereais (OUYANG et al., 2007) e apresenta sintenia com os genomas do milho e do trigo (MOORE et al., 1995). A disponibilidade de protocolos para a transformação genética mediada por *A. tumefaciens* (UPADHYAYA et al., 2000) possibilita estudos fisiológicos, genéticos e moleculares, fundamentais para o entendimento dos mais variados processos biológicos.

1.4 GENES ASR (ABA, STRESS AND RIPENING)

Genes ASR (do ingles *absiscic acid*, *stress and ripening*) foram inicialmente descritos em tomate (IUSEM et al., 1993) e tem sido identificados exclusivamente em plantas vasculares, muito embora estejam ausentes na planta modelo *A. thaliana* (CARRARI; FERNIE; IUSEM, 2004).

Suas funções têm sido relacionadas ao desenvolvimento dos frutos (CAKIR et al., 2003; CHEN et al., 2011), bem como à resposta da planta a estresses abióticos (ARENHART et al., 2013a; DAI et al., 2011; HSU et al., 2011; HU et al., 2013; JHA et al., 2012; JOO et al., 2013a, 2013b; KALIFA et al., 2004a; KIM et al., 2009; LIU et al., 2012; YANG et al., 2005) e bióticos (LIU et al., 2010).

Uma característica pertinente à proteínas ASR é a presença de dois domínios altamente conservados (YANG et al., 2008). O primeiro é composto por seis a sete resíduos de histidina na região amino-terminal com atividade de ligação a DNA dependente de zinco (ÇAKIR et al., 2003; KALIFA et al., 2004a). O segundo domínio compreende a maior parte da região carboxi-terminal, onde também se identifica o sinal de localização nuclear, sendo esta região denominada de domínio WDS (do ingles, *water*, *deficit*, *stress*). Na figura 1 pode ser observado o alinhamento das proteínas ASR de arroz, com destaque para o domínio WDS.

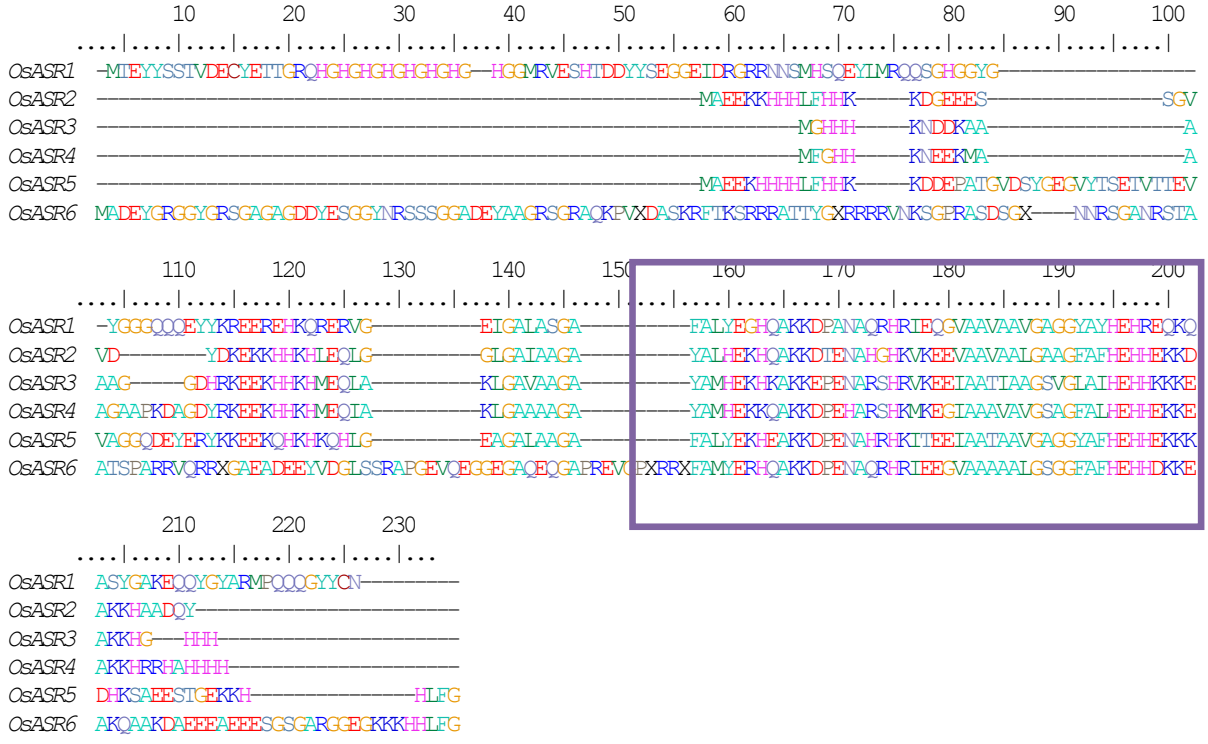

Figura 1. Alinhamento das seqüências de aminoácidos das seis proteínas da família ASR de arroz. Em destaque, o domínio WDS conservado entre os membros (conforme ARENHART et al., 2008).

A proteína ASR1 de tomate é eminentemente desestruturada (*unfolded*) e monomérica na ausência de zinco, sendo o metal fundamental para a formação de homodímeros e maior ordenamento (*fold*) na estrutura da proteína (GOLDGUR et al., 2007; ROM et al., 2006). Por outro lado, a proteína ASR5 de arroz não é capaz de formar homodímeros, muito embora a ligação de zinco também tenha sido confirmada (ARENHART et al., 2014).

Atuando tanto como chaperonas (KONRAD; BAR-ZVI, 2008) quanto como fatores de transcrição (ARENHART et al., 2014; RICARDI et al., 2014), essa família de proteínas desempenha papel na resposta das plantas aos mais variados estímulos ambientais. Quando superexpressas em Arabidopsis, proteínas ASR de lírio foram capazes de conferir menor suscetibilidade à seca, bem como aumentar o índice de germinação de sementes em concentrações inibitórias de manitol e sal, indicando uma conservação dos mecanismos *downstream* à proteína (YANG et al., 2005).

Análises *in silico* revelaram seis cópias de genes *ASR* no genoma do arroz, estando dispersas em diferentes cromossomos (Frankel et al., 2006) (Tabela 1).

Tabela 1. Localização, tamanho do íntron (em pb), e da proteína (em aminoácidos) dos genes *ASR* em arroz. Dados extraídos e modificados de Frankel *et al* (2006).

	Cromossomo	**Tamanho do íntron**	**Tamanho da proteína (aa)**	**ESTs**	**Em *tandem* com**
Arroz					
ASR1	II	*splicing*	63/71/91/105	sim	
ASR2	I	440	182	sim	
ASR3	I	131	105	sim	*ASR4*
ASR4	I	131	96	sim	*ASR3*
ASR5	XI	119	138	sim	
ASR6	IV	84	229	sim	

(aa) = aminoácidos

Splicing = diferentes formas de transcritos

ESTs = Expressed sequence tags

Em arroz, proteínas ASR foram inicialmente identificadas em biblioteca de cDNA de plantas submetidas a altas concentracões de sal e, posteriormente, também caracterizadas como sendo responsivas a ABA e manitol

(VAIDYANATHAN; KURUVILLA; THOMAS, 1999). Seu possível vínculo na resposta a estímulo hormonal foi previamente sugerida (TAKASAKI et al., 2008), bem como seu envolvimento na regulação de genes relacionados à fotossíntese (ARENHART et al., 2013). Em levedura (*Saccharomyces cerevisiae*), a superexpressão de proteínas ASR de arroz foi capaz de aliviar a produção de espécies reativas de oxigenio (EROs) causadas por estresse oxidativo (KIM; KIM; YOON, 2012). Plantas transgênicas de arroz superexpressando proteinas ASR foram mais tolerantes ao frio (JOO et al., 2013a; KIM et al., 2009) e seca (JOO et al., 2013b), quando comparadas à plantas não transgênicas. Recentemente, fatores de transcrição do tipo ASR foram identificados como componentes fundamentais na resposta a estresse por altas concentrações de alumínio em plantas de arroz. O referido estudo indicou que a expressão dos membros dessa família em arroz depende do tecido ou estímulo específico. A proteína ASR5 é a mais expressa em raízes e, acredita-se, ser componente fundamental no mecanismo de resposta ao estresse decorrente de altas concentrações de alumínio (ARENHART et al., 2013).

1.5 miRNAs E O PAPEL NA RESPOSTA A ESTRESSES ABIÓTICOS E BIÓTICOS

MicroRNAs (miRNAs) é uma classe de pequenos RNAs não codificantes, processados a partir de um grampo precursor, de maneira precisa, e cuja função é reprimir o mRNA alvo através de clivagem ou inibição traducional durante a regulação da expressão gênica (CHEN, 2009; JONES-RHOADES; BARTEL, 2004; JONES-RHOADES; BARTEL; BARTEL, 2006;). Estimativas indicam que 1-4% dos genes no genoma humano codificam miRNAs e que um único miRNA é capaz de regular até 200 mRNAs (ESQUELA-KERSCHER; SLACK, 2006). Fatores de transcrição têm sido identificados como ativadores ou repressores de miRNAs em plantas. Um exemplo é o mecanismo de sinalização PHR1-miR399-PHO2, envolvido na homeostase de fósforo (BARI et al., 2006). PHR1 (*Phosphate Starvation Response* 1) controla a expressão do miR399. Quando fósforo se torna um recurso limitante, PHR1 é ativado e induz a expressão do miR399, reprimindo a expressão de PHO2 (uma enzima de conjugação de ubiquitina tipo E2), a qual regula negativamente a captação de fósforo.

Em Arabidopsis, miRNAs mostraram-se essenciais para o correto

desenvolvimento da raiz (CARLSBECKER et al., 2010) e a relação entre fatores de transcrição e miRNAs foi descrita na rota de sinalização de auxinas no desenvolvimento de raízes adventícias (GUTIERREZ et al., 2009). Em plantas, mutações em genes envolvidos na biogênese de miRNAs e no seu mecanismo de regulação afetam o desenvolvimento (CHEN, 2009; RAMACHANDRAN; CHEN, 2008; XIE; KHANNA; RUAN, 2010). Em mutantes de arroz, insensíveis à auxina, um circuito de *feedback* entre a família miR167 e OsARF6 (*auxin responsive fator* 6) tem sido proposto como um importante *loop* regulatório na sinalização do fitohormônio auxina ou no desenvolvimento da raiz (MENG et al., 2009).

Muitos resultados também indicam que os miRNAs estão envolvidos na regulação de uma variedade de genes em resposta a estresses abióticos e bióticos. Um miRNA é o regulador chave do metabolismo do sulfato, em plantas com deficiência do metal (JONES-RHOADES; BARTEL, 2004). O mesmo fenômeno foi caracterizado em resposta à deficiência de fosfato (FUJII et al., 2005). Durante a limitação de cobre, miRNAs são induzidos e reprimem seu alvo regulatório, mantendo o controle da homeostase (YAMASAKI et al., 2007). Diversos miRNAs apresentaram os níveis de expressão aumentados em condições limitantes de ferro, indicando seu possível papel na adaptação das plantas à deficiência do metal (KONG; YANG, 2010).

Em um estudo pioneiro, o papel regulatório dos miRNAs na resposta a alumínio em arroz também foi sugerido. Raízes de cultivares tolerante e sensível foram expostas a altas concentrações do metal e miRNAs de diferentes famílias foram analisados. Os possíveis genes alvos identificados sugerem que os miRNAS de arroz estão envolvidos no controle de várias rotas metabólicas em resposta à exposição ao metal (LIMA et al., 2011).

O miR393 de Arabidopsis foi o primeiro pequeno RNA implicado na PTI bacteriana (PTI – do inglês, *PAMP-triggered immunity*, imunidade desencadeada por PAMP; PAMP – do inglês, *pathogen-associated molecular patterns*, padrão molecular associado ao patógeno – NAVARRO et al., 2006). A transcrição do MIR393 é induzida pelo peptídeo derivado da flagelina (chamado de flagelina 22) e degrada o mRNA da proteína F-box receptora da auxina (TIR1 – do inglês, transport inhibitor response 1) e proteínas relacionadas. Em outro exemplo, foi observado que o miR825 de Arabidopsis tem como alvo tres potenciais reguladores positivos da PTI (EULALIO et al., 2007; FAHLGREN et al., 2007).

Apesar de numerosos estudos demonstrarem a importância dos miRNAs como mediadores na regulação da expressão gênica, o mecanismo da regulação dos próprios miRNAs ainda é pouco conhecido. Estudos indicam que os genes MIR de plantas são transcritos pela RNA polimerase II (MEGRAW et al., 2006; XIE et al., 2005; ZHOU et al., 2007), situação similar ao que ocorre em animais (CAI; HAGEDORN; CULLEN, 2004; LEE et al., 2004). Com o objetivo de identificar e analisar a região promotora dos genes MIR em Arabidopsis, Zhao et al. (ZHAO; ZHANG; LI, 2013) realizaram um experimento de imunoprecipitação da enzima RNA polimerasedo tipo II, seguido por análise de microarranjo (ChIP-chip). Com base nos motivos de ligação da proteína ao DNA, foram preditos os sítios de início da transcrição e as regiões proximais dos promotores de 167 genes codificantes de miRNAs.

Apesar do progresso obtido em anos recentes, a descoberta de proteínas envolvidas no controle da expressão dos miRNAs, bem como a identificação de cis-elementos dos promotores de genes MIR é fundamental para um melhor entendimento das redes regulatórias nas quais os miRNAs possuem papel crucial.

1.6 ESPECTROMETRIA DE MASSA

O emprego de estratégias quantitativas para análise em larga escala de transcritos tem esclarecido aspectos relacionados tanto ao desenvolvimento quanto a fisiologia de plantas, porém, reações enzimáticas e rotas de sinalização dependem da atividade de proteínas, fonte de informação não contemplada por tais técnicas.

O balanço entre a síntese e a degradação de proteínas determina sua abundância e esse processo é independente do controle transcricional (PIQUES et al., 2009). Além disso, modificações pós-traducionais, isoformas e variantes de *splice* não são capturados pela mera análise da quantidade de transcritos.

Porém, modernas técnicas de espectrometria de massa possibilitam o estudo da complexidade do proteoma. A análise quantitativa do conjunto de proteínas e a dinâmica de suas mudanças em várias condições de crescimento e estímulos tem se tornado uma abordagem amplamente utilizada, sendo a análise de milhares de proteínas uma ferramente extremamente valiosa (ARSOVA; ZAUBER; SCHULZE, 2012).

Recentemente, vários métodos para a análise quantitativa de proteomas tem sido desenvolvidos (BANTSCHEFF et al., 2007; DOMON; AEBERSOLD, 2010; SCHULZE; USADEL, 2010), dentre eles, a marcação de aminoácidos utilizando isótopos estáveis esta sendo empregada em pesquisas das mais variadas áreas de estudo (ENGELSBERGER et al., 2006; GOUW; KRIJGSVELD; HECK, 2010).

Experimentos de proteômica quantitativa tem aprofundado o conhecimento sobre variados aspectos da biologia de organelas, regulação do crescimento e também sinalização (SCHULZE; USADEL, 2010). Por exemplo, mudanças na abundância de proteínas foram monitoradas em resposta ao calor (PALMBLAD; MILLS; BINDSCHEDLER, 2008) e durante a senescência das folhas (HEBELER et al., 2008).

Dessa forma, a técnica possui um grande potencial para identificar proteínas diferencialmente expressas nos momentos iniciais da resposta ao estresse por alumínio, com potencial para indentificar elementos chave na cascata de sinalização que ativa os mecanismos de adaptação da planta ao metal.

O excess de alumínio é limitante ao desenvolvimento das plantas, sendo o pH determinante na atividade biológica do metal. Dessa maneira, é a interação entre o baixo pH e o alumínio que determina a fitotoxicidade do metal. Compreender quem são e como atuam os elementos chave no processo de resposta a um ou ambos os estresses é fundamental. As proteínas ASR são importantes mediadores dessa resposta e, como tal, seu estudo é ferramenta indispensável para o entendimento da resposta da planta a esses estresses. Muito embora Arabidopsis não possua proteínas ASR, a identificação de genes envolvidos tanto na resposta ao pH quanto ao alumínio na planta modelo, demonstra uma conservacão dos mecanismos de sinalização tanto em monocotiledôneas quanto em eudicotiledôneas, validando seu uso em estudos genéticos e fisiológicos.

2. OBJETIVOS

2.1 OBJETIVO GERAL

O presente trabalho tem como objetivo analisar o papel das proteínas ASR na regulação de genes MIR, codificantes de miRNAs, bem como determinar seu possível papel na regulação do mecanismo de homeostase do pH em arroz. Além disso, este trabalho visa identificar proteínas potencialmente envolvidas nos mecanismos de defesa da planta em resposta ao metal alumínio.

2.1.1 Objetivos específicos:

1. Identificar miRNAs diferencialmente expressos em raízes de arroz (*Oryza sativa* cultivar Nipponbare) provenientes de plantas silenciadas para o gene *ASR5* e plantas não transformadas;
2. Determinar o padrão de expressão dos miRNAs identificados*;*
3. Identificar genes MIR potencialmente regulados pelas proteínas ASR5;
4. Avaliar o efeito do silenciamento do gene *ASR5* nas plantas transgênicas de arroz submetidas ao estresse provocado pelo baixo pH;
5. Comparar o perfil de expressão de proteínas diferencialmente expressas em plantas de *Arabidopsis thaliana* submetidas ao estresse pelo metal alumínio;
6. Identificar genes com potencial envolvimento no mecanismo de defesa da planta em resposta ao estresse por alumínio.

3. RESULTADOS E DISCUSSÃO

Os resultados e discussão serão apresentados em três capítulos. O capítulo 1 é dedicado à análise dos dados obtidos a partir do transcritoma de duas bibliotecas de microRNAseq de arroz, comparando o perfil de expressão de miRNAs de plantas silenciadas para o gene ASR5 (ASR5_RNAi) e plantas não transformadas. O capítulo 2 descreve o estudo das proteínas ASR na manutenção da homeostase do pH em plantas de arroz. O capítulo 3 é dedicado ao estudo da resposta inicial de plantas de *Arabidopsis thaliana* ao estresse por alumínio com o uso da técnica de espectrometria de massa.

REFERÊNCIAS BIBLIOGRÁFICAS

ARABIDOPSIS, T.; INITIATIVE, G. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. **Nature**, v. 408, n. 6814, p. 796–815, 14 dez. 2000.

ARENHART, R. A. **Análise funcional dos genes ASR - *Abscisic acid, Stress and Ripening* - de arroz (*Oryza sativa* L.) em resposta ao estresse por alumínio**. 2008. 94 f. Dissertação (Mestrado em Biologia Celular e Molecular). Departamento de Biologia Molecular e Biotecnologia. Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

ARENHART, R. A. et al. Involvement of ASR genes in aluminium tolerance mechanisms in rice. **Plant, cell & environment**, v. 36, n. 1, p. 52–67, jan. 2013.

ARENHART, R. A. et al. New insights into aluminum tolerance in rice: the ASR5 protein binds the STAR1 promoter and other aluminum-responsive genes. **Molecular plant**, v. 7, n. 4, p. 709–21, abr. 2014.

ARSOVA, B.; ZAUBER, H.; SCHULZE, W. X. Precision, proteome coverage, and dynamic range of Arabidopsis proteome profiling using ^{15}N metabolic labeling and label-free approaches. **Molecular & cellular proteomics: MCP**, v. 11, n. 9, p. 619–28, set. 2012.

BAJGUZ, A.; HAYAT, S. Effects of brassinosteroids on the plant responses to environmental stresses. **Plant physiology and biochemistry**, v. 47, n. 1, p. 1–8, jan. 2009.

BANTSCHEFF, M. et al. Quantitative mass spectrometry in proteomics: a critical review. **Analytical and bioanalytical chemistry**, v. 389, n. 4, p. 1017–31, out. 2007.

BAÑUELOS, M. A. et al. Inventory and Functional Characterization of the HAK Potassium Transporters of Rice. **Plant Physiology**, v. 130, n. October, p. 784–795, 2002.

BARCELO, J.; POSCHENRIEDER, C. Fast root growth responses , root exudates , and internal detoxification as clues to the mechanisms of aluminium toxicity and resistance : a review. **Environmental and Experimental Botany 48**, v. 48, p. 75–92, 2002.

BARI, R. et al. PHO2, MicroRNA399, and PHR1 Define a Phosphate-Signaling Pathway in Plants. **Plant Physiology**, v. 141, n. July, p. 988–999, 2006.

BOWN, A. W.; SHELP, B. J. The Metabolism and Functions of y-Aminobutyric Acid. **Plant physiology**, v. 1, n. 115, p. 1–5, 1997.

CAI, X.; HAGEDORN, C. H.; CULLEN, B. R. Human microRNAs are processed from capped , polyadenylated transcripts that can also function as mRNAs Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. **RNA**, v10, n. 12, p. 1957–1966, 2004.

CAKIR, B. et al. A grape ASR protein involved in sugar and abscisic acid signaling. **Plant Cell**, v. 15, n. 9, p. 2165–2180, 2003.

ÇAKIR, B. et al. A Grape ASR Protein Involved in Sugar and Abscisic Acid Signaling. **The Plant Cell**, v. 15, n. September, p. 2165–2180, 2003.

CAKMAK, I. Plant nutrition research: Priorities to meet human needs for food in sustainable ways. **Plant and Soil**, v. 247, p. 3–24, 2002.

CARLSBECKER, A. et al. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. **Nature**, v. 465, n. 7296, p. 316–21, 20 maio 2010.

CARRARI, F.; FERNIE, A. R.; IUSEM, N. D. Heard it through the grapevine? ABA and sugar cross-talk: the ASR story. **Trends in plant science**, v. 9, n. 2, p. 57–9, fev. 2004.

CHEN, J. et al. Molecular characterization of a strawberry FaASR gene in relation to fruit ripening. **PloS one**, v. 6, n. 9, p. e24649, jan. 2011.

CHEN, X. Small RNAs and their roles in plant development. **Annual review of cell and developmental biology**, v. 25, p. 21–44, jan. 2009.

CHENG, N.-H. et al. Arabidopsis monothiol glutaredoxin, AtGRXS17, is critical for temperature-dependent postembryonic growth and development via modulating auxin response. **The Journal of biological chemistry**, v. 286, n. 23, p. 20398–20406, 10 jun. 2011.

CHRISTENSEN, J. H. et al. Regional Climate Projections. In: **Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change**. p. 895–940. 2007.

CRAMER, G. R. et al. Effects of abiotic stress on plants: a systems biology perspective. **BMC plant biology**, v. 11, n. 1, p. 163, jan. 2011.

DAI, J.-R. et al. MpAsr encodes an intrinsically unstructured protein and enhances osmotic tolerance in transgenic Arabidopsis. **Plant cell reports**, v. 30, n. 7, p. 1219–1230, jul. 2011.

DELHAIZE, E.; MA, J. F.; RYAN, P. R. Transcriptional regulation of aluminium tolerance genes. **Trends in plant science**, v. 17, n. 6, p. 341–8, jun. 2012.

DOMON, B.; AEBERSOLD, R. Options and considerations when selecting a quantitative proteomics strategy. **Nature biotechnology**, v. 28, n. 7, p. 710–21, jul. 2010.

DRISCOLL, C. T.; SCHECHER, W. D. The chemistry of aluminum in the environment. **Environmental geochemistry and health**, v. 12, n. 1-2, p. 28–49, mar. 1990.

EMBRAPA. **Cultivo do Arroz Irrigado no Brasil**. Disponível em: <http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Arroz/ArrozIrrigadoBrasil/index.htm>. Acesso em: 15 maio. 2014.

EMBRAPA. Centro Nacional de Pesquisa de Solos (Rio de Janeiro, RJ). **Sistema brasileiro de classificação de solos**. 2. ed. Rio de Janeiro: EMBRAPA-SPI, 2006.

ENGELSBERGER, W. R. et al. Metabolic labeling of plant cell cultures with $K^{15}NO3$ as a tool for quantitative analysis of proteins and metabolites. **Plant methods**, v. 2, n. 3, p. 14, jan. 2006.

ESQUELA-KERSCHER, A.; SLACK, F. J. Oncomirs - microRNAs with a role in cancer. **Nature reviews. Cancer**, v. 6, n. 4, p. 259–69, abr. 2006.

EULALIO A. et al. Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. **Genes and Development***,* v. 21, n. 20, p. 2558–2570. out. 2007.

FAHLGREN, N. et al. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of *MIRNA* genes. **PLoS ONE**, v. 2, n. 2, p. e219, fev. 2007.

FAMOSO, A. N. et al. Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms. **Plant physiology**, v. 153, n. 4, p. 1678–1691, ago. 2010.

FUCHS, I. et al. Rice K^+ uptake channel OsAKT1 is sensitive to salt stress. **Planta**, v. 221, n. 2, p. 212–21, maio 2005.

FUJII, H. et al. A miRNA involved in phosphate-starvation response in Arabidopsis. **Current biology**, v. 15, n. 22, p. 2038–2043, 22 nov. 2005.

GIBALA, M. et al. The lack of mitochondrial AtFtsH4 protease alters Arabidopsis leaf morphology at the late stage of rosette development under short-day photoperiod. **The Plant journal**, v. 59, n. 5, p. 685–699, set. 2009.

GOLDGUR, Y. et al. Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. **Plant physiology**, v. 143, n. 2, p. 617–28, fev. 2007.

GONZÁLEZ, R. M.; IUSEM, N. D. Twenty years of research on Asr (ABA-stress-ripening) genes and proteins. **Planta**, v. 239, n. 5, p. 941–9, maio 2014.

GOUW, J. W.; KRIJGSVELD, J.; HECK, A. J. R. Quantitative proteomics by metabolic labeling of model organisms. **Molecular & cellular proteomics: MCP**, v. 9, n. 1, p. 11–24, jan. 2010.

GREVENSTUK, T.; ROMANO, A. Aluminium speciation and internal detoxification mechanisms in plants: where do we stand? **Metallomics: integrated biometal science**, v. 5, n. 12, p. 1584–94, dez. 2013.

GUTIERREZ, L. et al. Phenotypic plasticity of adventitious rooting in Arabidopsis is controlled by complex regulation of auxin response factor transcripts and microRNA abundance. **The Plant cell**, v. 21, n. 10, p. 3119–32, out. 2009.

HEBELER, R. et al. Study of early leaf senescence in Arabidopsis thaliana by quantitative proteomics using reciprocal $^{14}N/^{15}N$ labeling and difference gel electrophoresis. **Molecular & cellular proteomics: MCP**, v. 7, n. 1, p. 108–20, jan. 2008.

HSU, Y.-F. et al. Lily ASR protein-conferred cold and freezing resistance in Arabidopsis. **Plant physiology and biochemistry**, v. 49, n. 9, p. 937–945, set. 2011.

HU, W. et al. TaASR1, a transcription factor gene in wheat, confers drought stress tolerance in transgenic tobacco. **Plant, cell & environment**, v. 36, n. 8, p. 1449-1464, ago. 2013.

HUANG, C. F. et al. A bacterial-type ABC transporter is involved in aluminum tolerance in rice. **The Plant cell**, v. 21, n. 2, p. 655–667, fev. 2009.

HUANG, C.-F. et al. A tonoplast-localized half-size ABC transporter is required for internal detoxification of aluminum in rice. **The Plant journal**, v. 69, n. 5, p. 857–867, mar. 2012.

HUANG, C.-F.; YAMAJI, N.; MA, J. F. Knockout of a bacterial-type ATP-binding cassette transporter gene, AtSTAR1, results in increased aluminum sensitivity in Arabidopsis. **Plant physiology**, v. 153, n. 4, p. 1669–1677, ago. 2010.

IBGE. Levantamento sistemático da produção agrícola. v. 27, n. 2, p. 1–84, 2014.

IUCHI, S. et al. Zinc finger protein STOP1 is critical for proton tolerance in Arabidopsis and coregulates a key gene in aluminum tolerance. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 23, p. 9900–9905, 2007.

IUSEM, N. D. et al. Tomato (Lycopersicon esculentum) Transcript Induced by Water Deficit and Ripening. **Plant physiology**, n. 102, p. 1353–1354, 1993.

JHA, B. et al. The SbASR-1 gene cloned from an extreme halophyte Salicornia brachiata enhances salt tolerance in transgenic tobacco. **Marine biotechnology**, v. 14, n. 6, p. 782–792, dez. 2012.

JIAN ZHENG S; FENG MA J; MATSUMOTO, H. High aluminum resistance in buckwheat. I. Al-induced specific secretion of oxalic acid from root tips. **Plant physiology**, v. 117, n. 3, p. 745–751, jul. 1998.

JONES-RHOADES, M. W.; BARTEL, D. P. Computational Identification of Plant MicroRNAs and Their Targets , Including a Stress-Induced miRNA. **Molecular Cell**, v. 14, p. 787–799, 2004.

JONES-RHOADES, M. W.; BARTEL, D. P.; BARTEL, B. MicroRNAS and their regulatory roles in plants. **Annual review of plant biology**, v. 57, p. 19–53, jan. 2006.

JOO, J. et al. Abiotic stress responsive rice ASR1 and ASR3 exhibit different tissue-dependent sugar and hormone-sensitivities. **Molecules and cells**, v. 35, n. 5, p. 421–435, maio 2013a.

JOO, J. et al. Rice ASR1 has function in abiotic stress tolerance during early growth stages of rice. **Journal of the Korean Society for Applied Biological Chemistry**, v. 56, n. 3, p. 349–352, 30 jun. 2013b.

KALIFA, Y. et al. The water- and salt-stress-regulated Asr1 (abscisic acid stress ripening) gene encodes a zinc-dependent DNA-binding protein. **Biochem J**, v. 381, n. Pt 2, p. 373–378, 2004a.

KALIFA, Y. et al. Over-expression of the water and salt stress-regulated Asr1 gene confers an increased salt tolerance. **Plant, Cell and Environment**, v. 27, n. 12, p. 1459–1468, dez. 2004b.

KIM, I.-S.; KIM, Y.-S.; YOON, H.-S. Rice ASR1 protein with reactive oxygen species scavenging and chaperone-like activities enhances acquired tolerance to abiotic stresses in Saccharomyces cerevisiae. **Molecules and cells**, v. 33, n. 3, p. 285–293, mar. 2012.

KIM, S. et al. Ectopic Expression of a Cold-Responsive OsAsr1 cDNA Gives Enhanced Cold Tolerance in Transgenic Rice Plants. **Molecules and cells**, v. 27, p. 449–458, 2009.

KOBAYASHI, Y. et al. Molecular and physiological analysis of Al^{3+} and H^{+} rhizotoxicities at moderately acidic conditions. **Plant physiology**, v. 163, n. 1, p. 180–192, set. 2013.

KOCHIAN, L. V et al. Mechanisms of metal resistance in plants : aluminum and heavy metals. **Plant and Soil**, v. 247, p. 109–119, 2002.

KONG, W. W.; YANG, Z. M. Identification of iron-deficiency responsive microRNA genes and cis-elements in Arabidopsis. **Plant physiology and biochemistry**, v. 48, n. 2-3, p. 153–159, 2010.

KONRAD, Z.; BAR-ZVI, D. Synergism between the chaperone-like activity of the stress regulated ASR1 protein and the osmolyte glycine-betaine. **Planta**, v. 227, n. 6, p. 1213–1219, 2008.

KOYAMA, H.; TODA, T.; HARA, T. Brief exposure to low-pH stress causes irreversible damage to the growing root in Arabidopsis thaliana: pectin-Ca interaction may play an important role in proton rhizotoxicity. **Journal of experimental botany**, v. 52, n. 355, p. 361–368, fev. 2001.

LARSEN, P. B. et al. ALS3 encodes a phloem-localized ABC transporter-like protein that is required for aluminum tolerance in Arabidopsis. **The Plant journal : for cell and molecular biology**, v. 41, n. 3, p. 353–363, fev. 2005.

LARSEN, P. B. et al. Arabidopsis ALS1 encodes a root tip and stele localized half type ABC transporter required for root growth in an aluminum toxic environment. **Planta**, v. 225, n. 6, p. 1447–1458, maio 2007.

LAUGIER, E. et al. Involvement of thioredoxin y2 in the preservation of leaf methionine sulfoxide reductase capacity and growth under high light. **Plant, cell & environment**, v. 36, n. 3, p. 670–682, mar. 2013.

LEE, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. **The EMBO journal**, v. 23, n. 20, p. 4051–4060, 13 out. 2004.

LI, Y.-F. et al. Transcriptome-wide identification of microRNA targets in rice. **The Plant journal**, v. 62, n. 5, p. 742–59, 1 jun. 2010.

LIMA, J. C. et al. Aluminum triggers broad changes in microRNA expression in rice roots. **Genetics and molecular research**, v. 10, n. 4, p. 2817–2832, jan. 2011.

LIU, H.-Y. et al. Characterization of a novel plantain Asr gene, MpAsr, that is regulated in response to infection of Fusarium oxysporum f. sp. cubense and abiotic stresses. **Journal of integrative plant biology**, v. 52, n. 3, p. 315–323, mar. 2010.

LIU, J. et al. Aluminum-activated citrate and malate transporters from the MATE and ALMT families function independently to confer Arabidopsis aluminum tolerance. **The Plant journal**, v. 57, n. 3, p. 389–399, fev. 2009.

LIU, J. et al. Isolation of an abscisic acid senescence and ripening inducible gene from litchi and functional characterization under water stress. **Planta**, n. 237, v. 4, p.1025-1036, abr. 2013.

MA, J. F. et al. Internal Detoxification Mechanism of Al in Hydrangea. **Plant Physiology**, v. 113, p. 1033–1039, 1997.

MA, J. F. et al. Response of rice to Al stress and identification of quantitative trait Loci for Al tolerance. **Plant & cell physiology**, v. 43, n. 6, p. 652–659, jun. 2002.

MA, J. F.; RYAN, P. R.; DELHAIZE, E. Aluminium tolerance in plants and the complexing role of organic acids. **Trends in plant science**, v. 6, n. 6, p. 273–278, jun. 2001.

MAGALHAES, J. V. Aluminum tolerance genes are conserved between monocots and dicots. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 26, p. 9749–9750, 27 jun. 2006.

MAHAJAN, S.; TUTEJA, N. Cold, salinity and drought stresses: an overview. **Archives of biochemistry and biophysics**, v. 444, n. 2, p. 139–158, 15 dez. 2005.

MARON, L. G. et al. Transcriptional profiling of aluminum toxicity and tolerance responses in maize roots. **New Phytologist**, v. 179, p. 116–128, 2008.

MCHALE, L. et al. Plant NBS-LRR proteins: adaptable guards. **Genome biology**, v. 7, n. 4, p. 212, jan. 2006.

MEEHL, G. A. et al. Global Climate Projections. In: **Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change**. p. 747–846. 2007.

MEGRAW, M. et al. MicroRNA promoter element discovery in Arabidopsis. **RNA**, v. 12, n. 9, p. 1612–1619, set. 2006.

MENG, Y. et al. Genome-wide survey of rice microRNAs and microRNA-target pairs in the root of a novel auxin-resistant mutant. **Planta**, v. 230, n. 5, p. 883–898, out. 2009.

MOORE, G. et al. Grasses, line up and form a circle. **Current Biology**, v. 5, n. 7, p. 737–739, 1995.

NAVARRO L. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. **Science**, v. 312, n. 5772, 436–339, abr. 2006.

OHYAMA, Y. et al. Characterization of AtSTOP1 orthologous genes in tobacco and other plant species. **Plant physiology**, v. 162, n. 4, p. 1937–1946, ago. 2013.

OUYANG, S. et al. The TIGR Rice Genome Annotation Resource: improvements and new features. **Nucleic acids research**, v. 35, n. Database issue, p. D883–887, jan. 2007.

PALMBLAD, M.; MILLS, D. J.; BINDSCHEDLER, L. V. Heat-shock response in Arabidopsis thaliana explored by multiplexed quantitative proteomics using differential metabolic labeling. **Journal of proteome research**, v. 7, n. 2, p. 780–5, fev. 2008.

PIQUES, M. et al. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. **Molecular systems biology**, v. 5, n. 314, p. 314, jan. 2009.

RAMACHANDRAN, V.; CHEN, X. Small RNA metabolism in Arabidopsis. **Trends in plant science**, v. 13, n. 7, p. 368–374, jul. 2008.

RICARDI, M. M. et al. Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription factor. **BMC plant biology**, v. 14, p. 29, jan. 2014.

ROM, S. et al. Mapping the DNA- and zinc-binding domains of ASR1 (abscisic acid stress ripening), an abiotic-stress regulated plant specific protein. **Biochimie**, v. 88, n. 6, p. 621–628, 2006.

ROUNDS, M. A; LARSEN, P. B. Aluminum-dependent root-growth inhibition in Arabidopsis results from AtATR-regulated cell-cycle arrest. **Current biology**, v. 18, n. 19, p. 1495–1500, 14 out. 2008.

RYAN, P. R. et al. A higher plant delta8 sphingolipid desaturase with a preference for (Z)-isomer formation confers aluminum tolerance to yeast and plants. **Plant physiology**, v. 144, n. 4, p. 1968–77, ago. 2007.

SAKANO, K. Revision of Biochemical pH-Stat: Involvement of Alternative Pathway Metabolisms. **Plant and Cell Physiology**, v. 39, n. 5, p. 467–473, 1 maio 1998.

SAWAKI, Y. et al. STOP1 Regulates Multiple Genes That Protect Arabidopsis from Proton and Aluminum Toxicities. **Plant Physiololgy**, v. 150, n. May, p. 281–294, 2009.

SCHULZE, W. X.; USADEL, B. Quantitation in mass-spectrometry-based proteomics. **Annual review of plant biology**, v. 61, p. 491–516, jan. 2010.

SPADONI, S. et al. Polygalacturonase-Inhibiting Protein Interacts with Pectin through a Binding Site Formed by Four Clustered Residues of Arginine and Lysine. **Plant Physiology**, v. 141, n. June, p. 557–564, 2006.

STOCKING, M. A. Tropical soils and food security: the next 50 years. **Science**, v. 302, n. 5649, p. 1356–9135, 21 nov. 2003.

SUNKAR, R. et al. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. **Trends in plant science**, v. 12, n. 7, p. 301–9, jul. 2007.

TAKASAKI, H. et al. Identification and characterization of a gibberellin-regulated protein, which is ASR5, in the basal region of rice leaf sheaths. **Molecular genetics and genomics**, v. 279, n. 4, p. 359–370, abr. 2008.

THIMM, O. et al. Mapman: a User-Driven Tool To Display Genomics Data Sets Onto Diagrams of Metabolic Pathways and Other Biological Processes. **The Plant Journal**, v. 37, n. 6, p. 914–939, mar. 2004.

UEXKÜLL, H. R. VON; MUTERT, E. Global extent , development and economic impact of acid soils. **Plant and Soil**, v. 171, p. 1–15, 1995.

UPADHYAYA, N. M. et al. Agrobacterium-mediated transformation of Australian rice cultivars Jarrah and Amaroo using modified promoters and selectable markers. **Australian Journal of Plant Physiology**, v. 27, n. 3, p. 201–210, 2000.

URANTOWKA, A. et al. Plant mitochondria contain at least two i-AAA-like complexes. **Plant molecular biology**, v. 59, n. 2, p. 239–252, set. 2005.

VAIDYANATHAN, R.; KURUVILLA, S.; THOMAS, G. Characterization and expression pattern of an abscisic acid and osmotic stress responsive gene from rice. **Plant Science**, v. 140, n. 1, p. 21–30, jan. 1999.

VELTHUIZEN, H. VAN et al. **Mapping biophysical factors that influence agricultural production and rural vulnerability**. Rome, 2007.

WANG, C. S. et al. Characterization of a desiccation-related protein in lily pollen during development and stress. **Plant & cell physiology**, v. 39, n. 12, p. 1307–1314, dez. 1998.

XIA, J. et al. Plasma membrane-localized transporter for aluminum in rice. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 43, p. 18381–1835, 26 out. 2010.

XIA, J.; YAMAJI, N.; MA, J. F. A plasma membrane-localized small peptide is involved in rice aluminum tolerance. **The Plant journal**, v. 76, n. 2, p. 345–55, out. 2013.

XIE, Z. et al. Expression of Arabidopsis MIRNA Genes. **Plant Physiology**, v. 138, n. August, p. 2145–2154, 2005.

XIE, Z.; KHANNA, K.; RUAN, S. Expression of microRNAs and its regulation in plants. **Seminars in cell & developmental biology**, v. 21, n. 8, p. 790–797, out. 2010.

YAMAJI, N. et al. A zinc finger transcription factor ART1 regulates multiple genes implicated in aluminum tolerance in rice. **The Plant cell**, v. 21, n. 10, p. 3339–3349, out. 2009.

YAMASAKI, H. et al. Regulation of copper homeostasis by micro-RNA in Arabidopsis. **The Journal of biological chemistry**, v. 282, n. 22, p. 16369–16378, 1 jun. 2007.

YANG, C. Y. et al. A Lily ASR protein involves abscisic acid signaling and confers drought and salt resistance in Arabidopsis. **Plant Physiology**, v. 139, n. 2, p. 836–846, 2005.

YANG, C.-Y. et al. The LLA23 protein translocates into nuclei shortly before desiccation in developing pollen grains and regulates gene expression in Arabidopsis. **Protoplasma**, v. 233, n. 3-4, p. 241–54, nov. 2008.

YOKOSHO, K.; YAMAJI, N.; MA, J. F. An Al-inducible MATE gene is involved in external detoxification of Al in rice. **The Plant journal**, v. 68, n. 6, p. 1061–1069, dez. 2011.

ZHAO, X.; ZHANG, H.; LI, L. Identification and analysis of the proximal promoters of microRNA genes in Arabidopsis. **Genomics**, v. 101, n. 3, p. 187–194, mar. 2013.

ZHOU, X. et al. Characterization and identification of microRNA core promoters in four model species. **PLoS computational biology**, v. 3, n. 3, p. e37, 9 mar. 2007.

ZHU, X. F. et al. XTH31, encoding an in vitro XEH/XET-active enzyme, regulates aluminum sensitivity by modulating in vivo XET action, cell wall xyloglucan content, and aluminum binding capacity in Arabidopsis. **The Plant cell**, v. 24, n. 11, p. 4731–4747, nov. 2012.

ZHU, Y. et al. Adaptation of plasma membrane H(+)-ATPase of rice roots to low pH as related to ammonium nutrition. **Plant, cell & environment**, v. 32, n. 10, p. 1428–1440, out. 2009.

ANEXOS: OUTROS ARTIGOS CIENTÍFICOS PRODUZIDOS DURANTE O PERÍODO DE DOUTORADO

Genetics and Molecular Biology, 35, 1 (suppl), 233-246 (2012)

www.sbg.org.br

Research Article

Identification and *in silico* characterization of soybean trihelix-GT and bHLH transcription factors involved in stress responses

Marina Borges Osorio*, Lauro Bücker-Neto*, Graciela Castilhos, Andreia Carina Turchetto-Zolet, Beatriz Wiebke-Strohm, Maria Helena Bodanese-Zanettini and Márcia Margis-Pinheiro

Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Abstract

Environmental stresses caused by either abiotic or biotic factors greatly affect agriculture. As for soybean [*Glycine max* (L.) Merril], one of the most important crop species in the world, the situation is not different. In order to deal with these stresses, plants have evolved a variety of sophisticated molecular mechanisms, to which the transcriptional regulation of target-genes by transcription factors is crucial. Even though the involvement of several transcription factor families has been widely reported in stress response, there still is a lot to be uncovered, especially in soybean. Therefore, the objective of this study was to investigate the role of bHLH and trihelix-GT transcription factors in soybean responses to environmental stresses. Gene annotation, data mining for stress response, and phylogenetic analysis of members from both families are presented herein. At least 45 bHLH (from subgroup 25) and 63 trihelix-GT putative genes reside in the soybean genome. Among them, at least 14 bHLH and 11 trihelix-GT seem to be involved in responses to abiotic/biotic stresses. Phylogenetic analysis successfully clustered these with members from other plant species. Nevertheless, bHLH and trihelix-GT genes encompass almost three times more members in soybean than in *Arabidopsis* or rice, with many of these grouping into new clades with no apparent near orthologs in the other analyzed species. Our results represent an important step towards unraveling the functional roles of plant bHLH and trihelix-GT transcription factors in response to environmental cues.

Key words: drought, gene expression, *Glycine max*, phylogeny, plant-microbe interactions.

Introduction

Soybean [*Glycine max* (L.) Merril] is one of the most important crop species in the world. It is widely used for both human and animal consumption due to the high protein and oil contents of its grains. More recently, the potential for using soybean oil in renewable fuel production has also emerged (Programa Nacional de Produção e Uso de Biodiesel). Since it belongs to the Fabaceae family, soybean also takes part in the process of organic nitrogen fertilizer production through its symbiotic association with nitrogen-fixing bacteria (Gepts *et al.*, 2005). Currently, soybean producers are primarily concerned with losses caused by drought stress, Asian Soybean Rust (ASR, caused by the fungus *Phakopsora pachyrhizi*) and soybean cyst nematode (SCN, caused by *Heterodera glycines*) (EMBRAPA, 2007). Furthermore, the genetic variability found in soybean germplasm for those characteristics is restricted, which increases the vulnerability of this species to environmental stresses (Priolli *et al.*, 2002; Miles *et al.*, 2006).

As sessile organisms, higher plants are continuously exposed to a great variety of environmental stimuli. Because their survival depends on the ability to cope with those stimuli, plants have evolved a variety of sophisticated molecular mechanisms in response to environmental stresses. These generally involve alterations in gene expression, leading to changes in plant physiology, metabolism and developmental activities. Whether caused by abiotic (such as drought, salt and cold) or biotic factors (such as pathogens and insects), environmental stresses have serious adverse effects on agriculture. Therefore, a thorough understanding of the molecular mechanisms involved in plant stress tolerance has become pivotal for the development of new strategies and technologies related to the increasing demand on agricultural production (Rao *et al.*, 2006; Yoshioda and Shinozaki, 2009).

Upon stimuli perception, responses of plants to environmental stresses comprise the activation of a multitude of interconnected signaling pathways (Singh *et al.*, 2002). The phytohormones abscisic acid (ABA), ethylene (ET), jasmonic acid (JA) and salicylic acid (SA), aside from reactive

Send correspondence to Márcia Pinheiro Margis. Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: marcia.margis@ufrgs.br.
*These authors contributed equally to the work.

oxygen species (ROS), are known to act as messenger molecules that trigger specific (but at times overlapping) pathways of this complex network, leading to the accumulation of stress-related gene products (Yoshioda and Shinozaki, 2009). Besides, a great number of studies have highlighted the importance of the transcriptional regulation of target-genes through transcription factors in plant responses to environmental stresses (Zhou *et al.*, 2008; Chen *et al.*, 2009; Zhang *et al.*, 2009). Transcription factors act by binding to *cis*-elements in the promoter regions of target-genes, thereby activating or repressing their expression. Transcriptional reprogramming is known to result in both spatially and temporally altered expression patterns of stress-related genes. Thus, transcription factors are key players in fine-tuning stress responses at the molecular level (Singh *et al.*, 2002; Eulgem, 2005).

A large part of a plant's genome is devoted to transcription. With the recent completion of the soybean genome sequencing and assembly, a comparative analysis of putative transcription factor-encoding genes found in both soybean and the model dicot *Arabidopsis thaliana* can be performed. In the leguminous plant (whose genome is six times larger than that of *A. thaliana*), over 5,600 transcription factors were identified, these corresponding to about 12% of the predicted protein-coding loci (Schmutz *et al.*, 2010). In contrast, in the model plant the total number of transcription factors (~2,300) comprises only up to 7% of the predicted protein-coding loci (Singh *et al.*, 2002). The overall distribution of these genes among known transcription-factor families is similar among the two genomes, although some families are relatively sparser or more abundant in soybean. Thus, even though the *A. thaliana* genome often serves general comparisons, differences in biological function between species might occur (Schmutz *et al.*, 2010).

Basic helix-loop-helix (bHLH) proteins constitute one of the largest families of transcription factors. They are found in all three eukaryotic kingdoms and are involved in a myriad of regulatory processes. Members of this family share the bHLH signature domain, which consists of ~60 amino acids comprising two distinct regions, a basic stretch at the N-terminus consisting of ~15 amino-acids involved in DNA binding, and a C-terminal region of ~40 amino-acids composed of two amphipathic α-helices, mainly consisting of hydrophobic residues linked by a variable loop (the "helix-loop-helix" region). This region is responsible for promoting protein-protein interactions through the formation of homo- and hetero-dimeric complexes (Toledo-Ortiz *et al.*, 2003; Carretero-Paulet *et al.*, 2010; Pires and Dolan, 2010). The Lc protein from *Zea mays,* reported as a transcriptional activator in the anthocyanin biosynthetic pathway (Ludwig *et al.*, 1989), was the first plant bHLH member identified. The involvement of bHLH members in plant developmental processes (Szecsi *et al.*, 2006; Menand *et al.*, 2007), light perception (Liu *et al.*, 2008), iron and phosphate homeostasis (Yi *et al.*, 2005; Long *et al.*, 2010; Zheng *et al.*, 2010), and phytohormone signalling pathways (Abe *et al.*, 1997; Friedrichsen *et al.*, 2002; Lorenzo *et al.*, 2004; Anderson *et al.*, 2004; Fernandez-Calvo *et al.*, 2011; Hiruma *et al.*, 2011; Seo *et al.*, 2011) has also been reported. In fact, *Arabidopsis* MYC2 is to date the most extensively characterized plant bHLH transcription factor, and it seems to be a global regulator of hormone signalling. MYC2 has been described as an activator of ABA-mediated drought stress-response (Abe *et al.*, 1997, 2003). It also regulates JA/ET-induced genes, either as an activator in response to wounding, or as a suppressor in pathogen responses (Anderson *et al.*, 2004; Lorenzo *et al.*, 2004; Hiruma *et al.*, 2011). In these cases, the activity of MYC2 is itself subject to regulation by JAZ proteins, in a SCF^{COI1} proteosome degradation – dependent pathway (Chini *et al.*, 2007). Additionally, MYC2 seems to form homo- and heterodimers with two other closely-related bHLH proteins (MYC3 and MYC4), and their interaction is essential for full regulation of JA responses in *Arabidopsis* (Fernandez-Calvo *et al.*, 2011).

Trihelix-GT factors constitute another family of plant-specific transcription factors. They are characterized by binding specificity for GT-elements present in the promoter region of many plant genes (Hiratsuka *et al.*, 1994; Nagano *et al.*, 2001) and are among the first transcription factors identified in plants (McCarty and Chory, 2000). They share one or two trihelix (helix – loop – helix – loop – helix) structures, each consisting of three putative α-helices, which are responsible for binding to DNA (Zhou, 1999). Dimerization of GT factors, or interaction between trihelix-GT and other transcription factors appear to play a major role in the regulatory function of this family (Zhou, 1999). In addition, recent studies demonstrated that post-translational modifications may occur in at least some GT-factors, as shown for *Arabidopsis* light-responsive GT-1 (Maréchal *et al.*, 1999; Nagata *et al.*, 2010). Members of the trihelix-GT family were first described as being involved in the regulation of light-responsive genes (Green *et al.*, 1987, 1988). Nevertheless, further studies in rice and *Arabidopsis* showed that some GT factors are not light-responsive at the transcriptional level (Dehesh *et al.*, 1990; Kuhn *et al.*, 1993). The involvement of this family in seed maturation (Gao *et al.*, 2009), control of flower morphogenesis (Griffith *et al.*, 1999; Brewer *et al.*, 2004; Li *et al.*, 2008), and response to environmental cues (O'Grady *et al.*, 2001; Park *et al.*, 2004; Wang *et al.*, 2004; Xie *et al.*, 2009; Fang *et al.*, 2010) has also been reported.

In recent years, a growing number of transcription factors belonging to families, such as AP2, NAC and WRKY, have been connected to the responses of soybean against environmental stresses (Zhang *et al.*, 2009; Pinheiro , 2009; Zhou , 2008). In addition, the involvement of two soybean trihelix-GT factors [*GmGT*-2A (Glyma04g39400) and *GmGT*-2B (Glyma10g30300)] in abiotic stress toler-

ance has recently been proposed, following heterologous expression in *Arabidopsis* (Xie , 2009). Nevertheless information regarding soybean bHLH and trihelix-GT members and their roles in this species remains scarce. In the present study we, therefore, aimed at identifying soybean bHLH- and trihelix-GT-encoding genes, as well as investigating their involvement in response to environmental stresses. Given the dimension of the bHLH family in plants (with more than 600 members in *Arabidopsis* divided into 32 groups), we decided to focus on a single monophyletic group (subfamily 25, Carretero-Paulet *et al.*, 2010), once we had found some interesting soybean candidates within the LGE Soybean Genome database (Nascimento *et al.*, 2012) that belong to this group. At least 45 bHLH (from subgroup 25) and 63 trihelix-GT putative genes reside in the soybean genome. Among these, at least 14 bHLH and 11 trihelix-GT seem to be involved in responses to abiotic/biotic stresses. A phylogenetic analysis allowed us to successfully cluster these genes with members of bHLH and trihelix-GT proteins from other plant species. All together, our results represent an important step towards understanding the molecular mechanisms by which soybean responds to environmental cues.

Material and Methods

Sequence identification and annotation

In order to identify putative soybean bHLH sequences, the TAIR (The *Arabidopsis* Information Resource) gene id from all 17 bHLH proteins belonging to subgroup 25 in *Arabidopsis* was used to search the soybean database in Phytozome and at JGI (Joint Genome Institute). Soybean peptide homologs for each *A. thaliana* sequence were identified from a BLASTP search with default parameters in Phytozome and redundant sequences were manually discarded. The protein sequences obtained were scanned for the existence of the bHLH domain using the SMART database. The software MEME (multiple EM for motif elicitation) version 4.4.0 was used for motif identification, using the following parameters: minimum and maximum motif width set to 6 and 50 amino acids, respectively, with any number of motif repetitions. Motif detection was restricted to a maximum of 10. Identified motifs were also compared with conserved compositions already described for bHLH sequences. In addition, the bHLH domain was manually delimited according to plant-specific boundaries, as determined by Toledo-Ortiz *et al.* (2003) and Carretero-Paulet *et al.* (2010). Classification of soybean sequences in subgroup 25 was accomplished by mismatch counting from the consensus established for *A. thaliana* (Carretero-Paulet *et al.*, 2010). Sequences with more than 8 mismatches in conserved positions were discarded. Moreover, no mismatches were allowed at residues H_9, E_{13} and R_{16} of the basic region, since these are crucial for DNA-binding activity, and a consensus among subgroup 25 sequences.

The identification of putative trihelix-GT protein sequences from soybean was accomplished as follows: the conserved trihelix sequence of previously reported soybean genes (O'Grady *et al.*, 2001; Xie *et al.*, 2009) along with motifs predicted for this family (Fang *et al.*, 2010), were blasted (TBLASTN) against the soybean genome in Phytozome. All homologous sequences with an E-value of less than 0.0001 were scanned for the existence of the trihelix domain using SMART (domains with less significant scores than default cut-offs were also analyzed). Motif identification and comparison with conserved trihelix-GT compositions were performed using MEME. Sequences that did not fit these criteria were removed from the analysis.

To determine the intron-exon organization of all *bHLH* and *trihelix-GT* genes, the full length coding sequences were aligned with the corresponding genomic sequences available on Phytozome. Intron-exon maps of the genes were drawn using Fancy Gene v1.4 software.

Gene expression data mining

Expression profiles of the identified bHLH and trihelix-GT sequences in both biotic and abiotic situations were obtained by mining the LGE Soybean Genome database. A "gene" search was carried out using Phytozome's gene model codes and each gene had its 5' and 3' untranslated regions verified in Gbrowse. Gene expression was confirmed by database searches in NCBI ESTs and LGE superSAGE stress experiments with soybean leaves infected with Asian soybean rust (accession PI 561356, resistant) *vs.* uninfected leaves, and soybean roots subjected to drought (cultivar BR16, susceptible / cultivar Embrapa-48, tolerant) *vs.* untreated roots from both cultivars.

Phylogenetic analysis

The phylogenetic analysis of plant trihelix-GT factors was performed using protein sequences from *A. thaliana*, *G. max*, *Medicago truncatula* and *Oryza sativa*. For plant bHLH transcription factors, protein sequences from *A. thaliana, G. max, O. sativa* and *Physcomitrella patens* were used. In both cases, multiple sequence alignments were conducted with full-length protein sequences using the CLUSTALW tool (Thompson *et al.*, 1994) implemented in MEGA ver. 4.0 (Tamura *et al.*, 2007). The phylogenetic analysis was performed by two different and independent approaches, viz. the neighbor-joining (NJ) and Bayesian methods. The NJ method was performed within MEGA v4.0. Molecular distances of the aligned sequences were calculated according to the p-distance parameter, with gaps and missing data treated as pairwise deletions. Branch points were tested for significance by bootstrapping with 1000 replications. Bayesian analysis was conducted in MrBayes 3.1.2 software (Huelsenbeck *et al.*, 2001; Ronquist and Huelsenbeck, 2003) with the mixed amino-acid substitution model + gamma + invariant sites. Two inde-

pendent runs of 5,000,000 generations each, with two Metropolis-coupled Monte Carlo Markov chains (MCMCMC) were run in parallel, each one 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 allcompat), and the posterior probability of clades and branch lengths. The unrooted phylogenetic trees of trihelix-GT and bHLH proteins were visualized and edited using the software FigTree ver. 1.3.1.

Results and Discussion

Identification and analysis of soybean bHLH-encoding genes

In the past few years several phylogenetic studies have emerged as attempts to perform the classification of bHLH proteins in plants (Heim *et al.*, 2003; Toledo-Ortiz *et al.*, 2003; Carretero-Paulet *et al.*, 2010; Pires and Dolan, 2010). Nevertheless, the number of proposed subfamilies varies considerably among these studies. In the present one, the classification suggested by Carretero-Paulet *et al.* (2010) proposing the division of plant bHLH transcription factors into 32 subfamilies was used, since it represents the most recent and comprehensive study, so far.

From the BLASTP search at Phytozome, using all 17 *Arabidopsis* bHLH protein sequences from subgroup 25, 67 non-redundant homolog peptides were identified in the soybean genome. Seven of these were removed from the analysis as they did not contain any bHLH domain. Another 15 sequences were discarded after mismatch counting performed with their aligned domains. Using MEME, two other highly conserved motifs (with E-values of less than 1.7 e^{-851}) were identified among the soybean subgroup 25 sequences. They are formed by residues right adjacent to the bHLH domain and had been previously reported (Heim *et al.*, 2003; Li *et al.*, 2006; Carretero-Paulet *et al.*, 2010; Pires and Dolan, 2010). General characteristics related to the 45 remaining putative soybean bHLH genes are shown in Table 1. Remarkably, members of this subgroup were found spread throughout the 20 soybean chromosomes, with protein sequences ranging from 165 to 691 amino acids. Among the 45 annotated ORFs, 42 presented corresponding ESTs, suggesting that they are expressed genes and not pseudogenes. A complete overview of the gene expression results obtained for this group is presented in Figure 1. Differential expression in at least one of the stress situations/experiments available in LGE database was detected for 14 ORFs, four of these were differentially expressed in more than one situation and three respond to both abiotic and biotic stresses.

Lately, a growing number of studies accessing the functional role of specific plant bHLH transcription factors have been reported (Friedrichsen *et al.*, 2002; Szécsi *et al.*,

Table 1 - Annotation of soybean bHLH (subgroup 25) encoding-genes.

Accession number in Phytozome	Chromosome	ORF (bp)	Expression confirmed by EST (GenBank Accession)
Glyma01g04610	1	795	BE021678.1
Glyma01g09400	1	1587	BU765737.1
Glyma01g39450	1	667	AW782148.1
Glyma02g13860	2	1539	BI786324.1
Glyma02g16110	2	861	AW460021.1
Glyma03g21770	3	1575	FK005566.1
Glyma03g29710	3	1203	BI427219.1
Glyma03g31510	3	879	BW666688.1
Glyma03g32740	3	1446	BM732402.1
Glyma04g01400	4	1293	CA853113.1
Glyma04g05090	4	855	FK457664.1
Glyma04g34660	4	732	FG990727.1
Glyma04g37690	4	1041	CA937888.1
Glyma05g01590	5	675	EV276804.1
Glyma05g35060	5	741	BE473364.1
Glyma05g38450	5	1029	BF325330.1
Glyma06g01430	6	1173	BU551063.1
Glyma06g17420	6	1050	FG995242.1
Glyma06g20000	6	810	CO978579.1
Glyma07g10310	7	498	BE347561.1
Glyma08g01210	8	942	FG994001.1
Glyma08g04660	8	528	-
Glyma08g46040	8	1761	BM885094.1
Glyma09g14380	9	1473	CA936197.1
Glyma09g31580	9	906	-
Glyma10g03690	10	852	BW657011.1
Glyma10g04890	10	1302	BI785116.1
Glyma10g12210	10	1074	CO978592.1
Glyma10g28290	10	2076	BW675573.1
Glyma10g30430	10	987	FG999826.1
Glyma11g05810	11	1146	GR843316.1
Glyma11g12450	11	1263	BU082612.1
Glyma12g04670	12	1215	BE661807.1
Glyma13g19250	13	1437	BQ741548.1
Glyma14g10180	14	1269	EV269688.1
Glyma15g33020	15	1428	BI699764.1
Glyma16g10620	16	1788	FK024158.1
Glyma17g08300	17	1098	CX708610.1
Glyma17g10290	17	690	FG993937.1
Glyma17g34010	17	807	-
Glyma18g32560	18	1743	BI317112.1
Glyma19g32570	19	1101	FG996268.1
Glyma19g34360	19	879	GR826097.1
Glyma20g22280	20	1281	BE658194.1
Glyma20g36770	20	999	BE474708.1

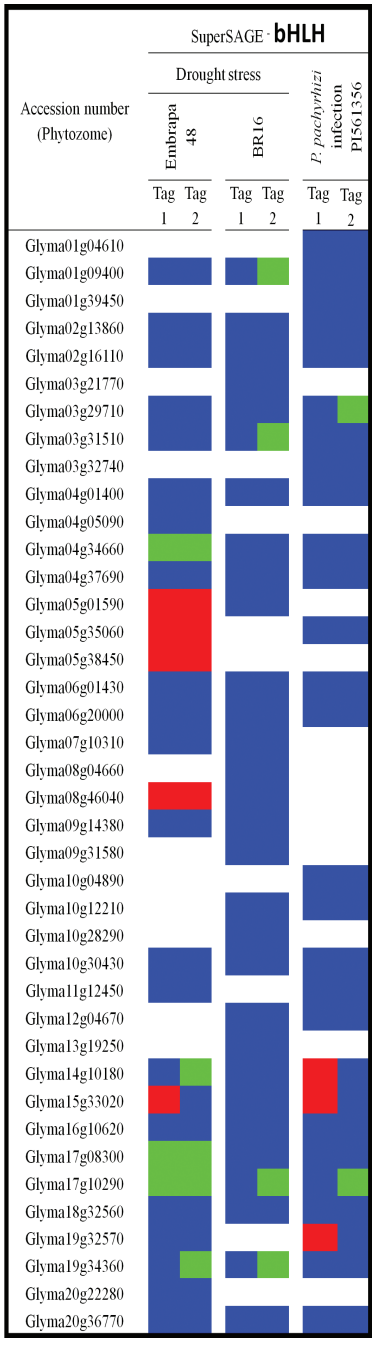

Figure 1 - Expression pattern of bHLH encoding-genes under drought stress and *P. pachyrhizi* infection. The expression data were obtained from superSAGE experiments available at www.lge.ibi.unicamp.br/soja/. Blocks indicate up-regulation (red), down-regulation (green), non-significant differences ($p > 0.05$) but expression detected (blue), and expression not detected (white). Contrasting expression might reflect detection of a single gene by different tags. Drought stress was carried out in roots from Embrapa-48 (tolerant cultivar) and BR 16 (susceptible cultivar). Soybean leaves from PI561356 (resistant genotype) were infected with *P. pachyrhizi*.

2006; Liu *et al.*, 2008; Chandler *et al.*, 2009; Todd *et al.*, 2010; Zheng *et al.*, 2010). Nevertheless, a deeper (and broader) functional characterization of this family, focusing on the connection of members/subgroups to the biological processes they control, remains to be done. A first step in this direction has been recently taken by Carretero-Paulet *et al.* (2010) and Pires and Dolan (2010), where comprehensive information relating both classification and function of previously characterized plant bHLH transcription factors was assembled. More specifically, information regarding the function of subgroup 25 members is still scarce and concerns *Arabidopsis* members only. An alternative transcript of At1g59640 (*ZCW*32/*BPE*) seems to be involved in the control of petal size, whereas its counterpart is expressed ubiquitously (Szécsi *et al.*, 2006). Furthermore, At4g34530 (*CIB*1) and At1g26260 (*CIB*5) were shown to interact with blue-light receptor CRY2 and promote floral initiation (Liu *et al.*, 2008). Of most interest for this study, is the redundant role of At1g18400 (*BEE*1, *Brassinosteroid Enhanced Expression*1), At4g36540 (*BEE*2) and At1g73830 (*BEE*3) in brassinosteroids (BRs)/ABA antagonistic cross-talk during cell elongation (Friedrichsen *et al.*, 2002). According to these authors, *BEE*1, 2 and 3 are early-response genes induced by BRs through the BRI1 receptor complex, and their expression is repressed by ABA through a yet unknown ABA receptor. Whether this pathway is also related to the ABA-dependent stress-responsive network, still requires further study. Moreover, Poppenberger *et al.* (2011) have demonstrated that At1g25330 (*CESTA*), a close homolog of *BEE*1 and *BEE*3 (Figure 2), is also involved in BR signaling, possibly by heterodimerization with its closest homologs. Remarkably, it has also been shown that lack of CESTA activity results in the misregulation of genes that are not only BR-responsive but also stress-responsive, such as *Arabidopsis ERD*5 (*Early Responsive to Dehydration* 5), *TTL4* (*Tetratricopetide-Repeat Thioredoxin-Like* 4), *WRKY18* and a putative LRR-disease resistance protein (Poppenberger *et al.*, 2011), further suggesting that these pathways might indeed share common features.

As an attempt to predict gene function of the annotated genes, a comparison of their amino-acid sequences with subgroup 25 bHLH protein sequences from three other model plant species was carried out. Indeed, representative members from diverse taxonomic groups (*P. patens*, bryophytes; *O. sativa*, monocotyledonous; and *A. thaliana*, dicotyledonous) were included in the phylogenetic analysis in order to access the evolutionary features of this subgroup. The results obtained from the phylogenetic analysis proved to be consistent, since the clades formed were highly supported by posteriori probabilities (Figure 2, on left) and bootstrap (data not shown) analyses. Unlike previous phylogenetic reconstructions of the bHLH family that used the bHLH domain only, this study presents a tree reconstructed from full-length protein sequences. This adds

accuracy and reliability to the tree resolution, since the short length of the bHLH domain (~60 amino-acids), along with its extremely high conservation within subgroups may compromise the reliability of the analysis (Amoutzias *et al.*, 2004).

Patterns of intron distribution among bHLH-encoding genes from diverse species were shown to be conserved within subgroups and provide another criterion in phylogenetic analysis (Li *et al.*, 2006; Carretero-Paulet *et al.*, 2010). In this study, the overall intron-exon organization of bHLH subfamily 25-encoding genes from soybean and other three species was established (Figure 2, on right). Among 89 sequences, the number of introns ranged from 1 (Pp1s270_17v6) to up to 12 (LOC_Os03g12940), and in many cases, phylogenetically related proteins exhibited a closely related gene structure, corroborating the clustering results.

Since it is a basal species among land plants, the moss *P. patens* was added to this classification in order to help infer about this group's ancestral state (Rensing *et al.*, 2008). Notably, all 12 members from *P. patens* grouped together into a clade, instead of grouping with the other plant species, indicating that the radiation within this subgroup has occurred independently in mosses and vascular plants, after the divergence of these taxonomic groups. The same result was obtained by Carretero-Paulet *et al.* (2010), even when a different method was applied [maximum likelihood (ML) analysis from bHLH-domain alignments]. Nevertheless, the chance that genes belonging to this subgroup might have independently evolved similar functions in both mosses and vascular plants should not be discarded, as suggested by Menand *et al.* (2007). In fact, while studying plant bHLH ancestry, Pires and Dolan (2010) concluded that the complex regulatory machinery that may be observed in modern plant lineages actually arose early in plant evolution.

The most striking feature that can be inferred from our phylogenetic analysis, which is in accordance with other previously published plant bHLH phylogenies (mentioned above), is the importance of gene duplication during the evolution of this family as a whole. Recurring events of single-gene duplications ("birth-and-death evolution"), combined with domain shuffling seem to rule bHLH evolution and diversification (Morgenstern and Atchley, 1999; Amoutzias *et al.*, 2004; Nei and Rooney, 2005). Furthermore, whole genome duplication (WGD) events also seem to have had an active effect (as seen in the outer clades in Figure 2, on the left), and this seems to be even more intense in the soybean genome. According to our results, the subgroup in question encompasses almost three times more members in soybean than in *Arabidopsis* or rice (Table 1), with many of these grouping into new clades with no apparent near orthologs in the other analyzed species (Figure 2, in gray on the left side). Indeed, soybean suffered from two WGD events with an impressive retention of homologous blocks (Schmutz *et al.*, 2010). Furthermore, specifically in the case of transcription factors (and other genes working in complex networks), duplications resulting from WGD events are vastly overretained, simply because they may be too costly to be removed, thus making functional redundancy a common feature among transcription factors, especially in plant species. Once retained, homologous duplicates might diverge in function or even subfunctionalize (Freeling, 2009), thus providing a source of evolutionary novelty in the form of new regulatory networks (Carretero-Paulet *et al.*, 2010).

With all that in mind, an integrated analysis of both the expression profile (Figure 1) and the phylogeny (Figure 2) presented herein provides a hint at the roles of subgroup 25 bHLH soybean genes. By focusing on soybean-near homologs shown in the tree (Figure 2 on left) we could see that for most of the paralogs whose expression has been detected, a divergent profile seems to prevail. An exception would be the cases of Glyma03g31510 and Glyma19g34360, which were both repressed during drought stress, with a broadly negative response in the latter, as its mRNA levels were down-regulated in both the susceptible and the tolerant cultivars analyzed. Moreover, the transcripts from Glyma19g32570 were up-regulated during ASR infection in the resistant genotype, whereas its counterpart Glyma03g29710 exhibited opposite differential expression. The near paralogs Glyma05g01590 and Glyma17g10290 also seem to be moving in different directions. Whereas the first seems to be up-regulated in response to fungal stress, the latter seems to be broadly down-regulated, in both susceptible and tolerant cultivars submitted to drought, as well as in *P. pachyrhizi*'s infection. Furthermore, while Glyma15g33020 seems to be positively involved in soybean defense against ASR and during drought stress in tolerant Embrapa-48 cultivar, its nearest paralog (Glyma09g14380) was not differentially expressed in any of the situations assessed, and their near homolog Glyma17g08300 seems to be negatively involved in drought stress responses, since it was down-regulated in the same cultivar. Whether the examples mentioned above reflect functional divergence or subfunctionalization among duplicate homologs still requires further analysis.

Even though comparison of soybean genes with their orthologs in other species (such as *Arabidopsis*) is a tentative approach, and as such needs to be performed carefully. In this context it would be interesting to address the function of *BEE* orthologs in soybean, so as to determine whether they are similar to their *Arabidopsis* counterparts, and whether they somehow connected to stress responses. In this respect, special attention should be given to Glyma05g35060, which clustered together with the *Arabidopsis* BR-responsive genes, and whose transcripts turned out to be up-regulated in Embrapa-48 tolerant cultivar in response to drought.

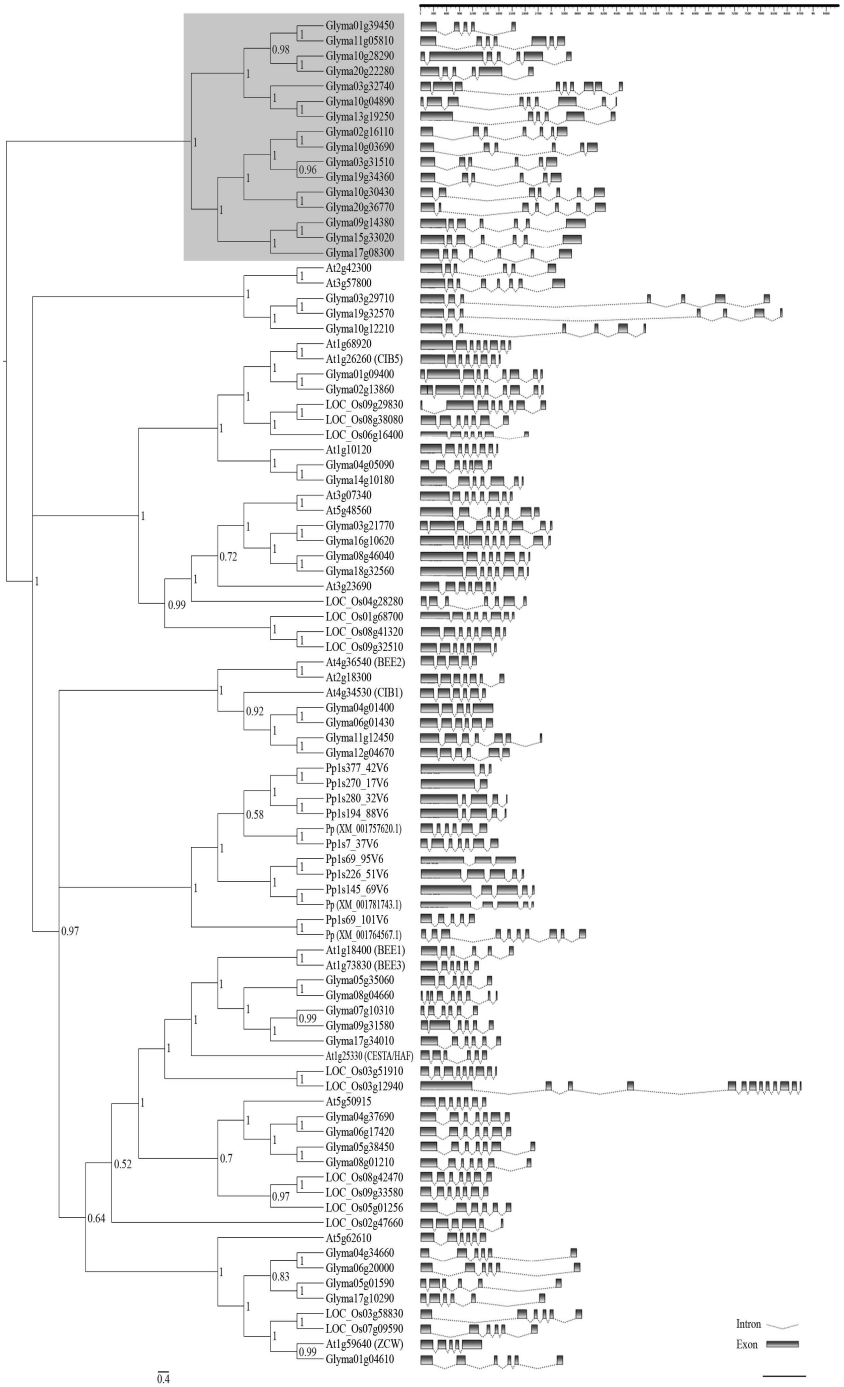

Figure 2 - Phylogenetic relationships among bHLH subgroup 25 members. The phylogenetic tree shown on the left comprises 89 plant bHLH protein sequences. The Bayesian analysis was conducted using Mr.Bayes v3.1.2, after alignment of full-length bHLH proteins from selected plant species by means of ClustalW. The unrooted cladogram was edited using Fig Tree v1.3.1 software. Nodal support is given by posteriori probability values shown next to the corresponding nodes. The scale bar indicates the estimated number of amino acid substitutions per site. The gray area denotes a specific soybean cluster. Previously reported bHLH genes were identified according to their accession/locus numbers, the other genes were designated according to their locus ID in Phytozome. *A. thaliana* (At); *G. max* (Glyma); *O. sativa* (LOC_Os) and *P. patens* (Pp). The graph on the right shows gene organization of full-length coding sequences from 89 plant bHLHs. Intron-exon maps were drawn using Fancy Gene v1.4 software, according to sequence data available in Phytozome.

Identification and analysis of soybean trihelix-GT encoding genes

The first isolated and described soybean GT-factor was *GmGT*-2 (Glyma02g09060), which binds to an element within the Aux28 promoter, and whose mRNA levels were down-regulated by light in a phytochrome-dependent manner (O'Grady *et al.*, 2001). In a global approach using massive EST analysis, Tian *et al.* (2004) identified 13 putative trihelix genes in the soybean genome. Two of these [*GmGT*-2A (Glyma04g39400) and *GmGT*-2B (Glyma10g30300)] were cloned and had their roles in abiotic stress tolerance described using transgenic *Arabidopsis* plants (Xie *et al.*, 2009). The current annotation analysis indicates the occurrence of at least 63 GT-like genes in the soybean genome. 56 of these had their expression confirmed in the NCBI databases (Table 2). Unfortunately, since information available in Phytozome is not yet definitive, full-length cDNAs were not obtained for most sequences, so only gene-models were considered for this analysis. The 63 soybean trihelix-GT genes encode proteins with lengths ranging from 201 to 885 amino acids, distributed across most of the soybean chromosomes, ex-

Table 2 - Annotation of soybean trihelix-GT encoding-genes.

Accession number in Phytozome (gene)	Chromosome	ORF (bp)	Expression confirmed by EST (GenBank Accession)
Glyma01g29760	1	819	BW682708.1
Glyma01g35370	1	834	GR826253.1
Glyma02g09050	2	1653	FG988995.1
Glyma02g09060 (GmGT-2)	2	1896	AF372498.1
Glyma03g18750	3	765	DB957166.1
Glyma03g34730	3	1368	FK016354.1
Glyma03g07590	3	822	-
Glyma03g34960	3	1617	BE555145.1
Glyma03g40610	3	1626	-
Glyma04g37020	4	2217	CO982525.1
Glyma04g39400 (GmGT-2A)	4	1335	AI900211.1
Glyma06g15500	6	1494	BW678214.1
Glyma06g17980	6	2655	EH258249.1
Glyma07g04790	7	1107	CO981809.1
Glyma07g09690	7	1083	BM731493.1
Glyma07g18320	7	876	-
Glyma08g05630	8	942	AW351117.1
Glyma08g28880	8	981	CO979268.1
Glyma09g01670	9	918	FK019218.1
Glyma09g19750	9	1155	BE659959.1
Glyma09g32130	9	1014	GR829369.1
Glyma09g38050	9	969	AI460860.1
Glyma10g36980	10	1335	BU765094.1
Glyma10g07490	10	1494	GD961953.1
Glyma10g34520	10	1374	BE820805.1
Glyma10g36950	10	1350	BU549085.1
Glyma10g36960	10	2004	BW666798.1
Glyma10g07730	10	1785	FG992486.1
Glyma10g30300 (GmGT-2B)	10	1746	CA953306.1
Glyma10g34610	10	1017	-
Glyma10g44620	10	978	GR827102.1
Glyma11g25570	11	1026	CO979922.1
Glyma11g37390	11	1125	BI317190.1
Glyma12g33850	12	924	CD415252.1
Glyma13g21350	13	1410	CX708572.1
Glyma13g26550	13	957	BI702330.1
Glyma13g30280	13	939	DB955747.1
Glyma13g21370	13	1464	CO981764.1
Glyma13g36650	13	921	CA800657.1
Glyma13g41550	13	1221	GD834531.1
Glyma13g43650	13	1014	EV282528.1
Glyma15g03850	15	1233	BF068981.1
Glyma15g08890	15	603	BM085616.1
Glyma15g12590	15	696	-
Glyma15g01730	15	1113	GD914877.1
Glyma16g01370	16	1113	CA801229.1
Glyma16g14040	16	801	CO980073.1
Glyma16g28240	16	1785	FK012336.1
Glyma16g28250	16	1395	BQ296282.1
Glyma16g28270	16	1332	-
Glyma17g13780	17	2433	BQ273464.1
Glyma18g01360 (GmGT-1)	18	1131	BG406222.1
Glyma18g43190	18	879	-
Glyma18g51790	18	990	BQ786728.1
Glyma19g37410	19	1359	GR845650.1
Glyma19g37660	19	1641	BF066376.1
Glyma19g43280	19	1803	FK019637.1
Glyma20g30630	20	1338	BG726775.1
Glyma20g30640	20	1935	BW679178.1
Glyma20g30650	20	1893	EH261764.1
Glyma20g32940	20	1572	FG988154.1
Glyma20g36680	20	1773	BE607585.1
Glyma20g39410	20	960	BI699475.1

cept for chromosomes 5 and 14. There is an average of 3.5 GT-factor-encoding genes per chromosome, with the highest number of 9 genes found in chromosome 10, whereas a single member was detected in chromosomes 12 and 17, respectively. Three genes (Glyma09g19750, Glyma10g34610 and Glyma20g30630) with incorrect gene model predictions were manually curated.

Mining the LGE gene expression superSAGE experiments revealed that 11 soybean trihelix-GT genes were differentially expressed in the abiotic/biotic conditions tested (Figure 3). In accordance with our analyses, five trihelix-GT genes were up-regulated under drought in the tolerant cultivar (Embrapa-48), whereas only two genes were down-regulated in this genotype. In the susceptible cultivar (BR16), Glyma10g34520 had its transcript levels increased in response to water deficit and the opposite situation occurred with Glyma10g36950. When plants were infected with *P. pachyrhizi*, only two genes displayed up-regulation of mRNA levels in response to biotic stress whereas two others seemed to be down-regulated. Interestingly, none of the soybean trihelix-GT previously reported as responsive to stress conditions and particularly to abiotic stress [GmGT-2A (Glyma04g39400) and GmGT-2B (Glyma10g30300)] were detected in the superSAGE experiments herein assessed. Divergence in experimental parameters and genotypes used might explain this unexpected result.

Transcript levels from Glyma01g35370 and Glyma20g30640 increased when plants were infected with ASR, while the opposite situation occurred with Glyma16g28240 and Glyma17g13780 mRNA levels. A rice GT-factor (Os*RML*1) was already reported to be upregulated in response to *Magnaporthe grisea* (Wang *et al.*, 2004), which corroborates a connection between pathogen attack and trihelix-GT gene regulation. It is also possible that Glyma01g35370 may be involved in plant responses to both abiotic and biotic stresses, since the gene expression profile was modulated during water deficit and *P. pachyrhizi* infection.

The superSAGE experiments suggested that, at least in some cases, the same gene has variable transcript levels in different cultivars and/or in response to different stresses or agents. For example, when water deficit was imposed on soybean plants, Glyma10g36950 was down-regulated in the susceptible (BR16) and the tolerant (Embrapa-48) cultivars, whereas its transcript levels did not change in response to ASR. In another case, Glyma09g38050 was up-regulated in response to drought stress in Embrapa-48, but no differences were detected in BR16. Furthermore, Glyma13g26550 was down-regulated in response to drought stress in the tolerant cultivar, whereas its expression in cultivar BR16 did not exhibit any alterations. In these cases, in addition to differential gene regulation, there may be other factors contributing to distinct regulatory function, such as post-translational modifications or variation in dimerization partners (Zhou, 1999).

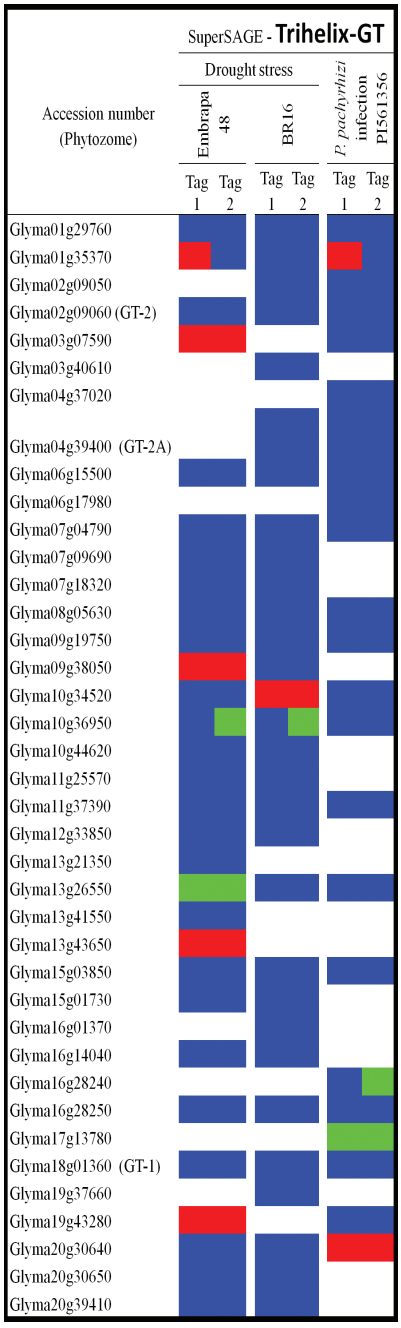

Figure 3 - Expression pattern of trihelix-GT encoding-genes under drought stress and *P. pachyrhizi* infection. The expression data were obtained from superSAGE experiments available at www.lge.ibi.unicamp.br/soja/. Blocks indicate up-regulation (red), down-regulation (green), non-significant differences ($p > 0.05$) but expression detected (blue), and expression not detected (white). Contrasting expression might reflect detection of a single gene by different tags. Drought stress was carried out in roots from Embrapa-48 (tolerant cultivar) and BR 16 (susceptible cultivar). Soybean leaves from PI561356 (resistant genotype) were infected with *P. pachyrhizi*.

Modifications in individual *cis*-regulatory elements on trihelix-GT promoter regions of duplicated genes might lead to the processes of transcriptional neofunctionalization or subfunctionalization (Haberer *et al.*, 2004), which may explain gene induction or repression without any counterpart response during the same stimuli. This seems to be the case for Glyma03g07590 and its nearest paralog Glyma01g29760, or for Glyma16g28240 and the phylogenetically related Glyma02g09050. Further studies focusing on identifying *cis*-elements, as well as performing promoter analyses to verify inducible expression patterns may clarify the involvement of duplicated genes in stress-related responses.

A previous study regarding the phylogenetic analysis encompassing *Arabidopsis* and rice GT factors (Fang *et al.*, 2010) showed that this family could be classified into three subfamilies (α , β and γ), with unique composition of predicted motifs. Unfortunately, these results were not reproduced in our analysis, even when full-length protein sequences (Figure 4) or the trihelix domains alone were aligned (data not shown). An exception occurred with subfamily γ, which had already been described as having low sequence similarity with the other reported GT factors. The introduction of soybean and *M. truncatula* sequences in the phylogeny might have affected the expected distribution within those subgroups. Besides, we also inserted into our tree the soybean gene AAK69274 described by Fang *et al.* (2010), which could neither be identified in the soybean genome nor detected in the expression database. According to our analysis, this unexpected result seems to indicate the

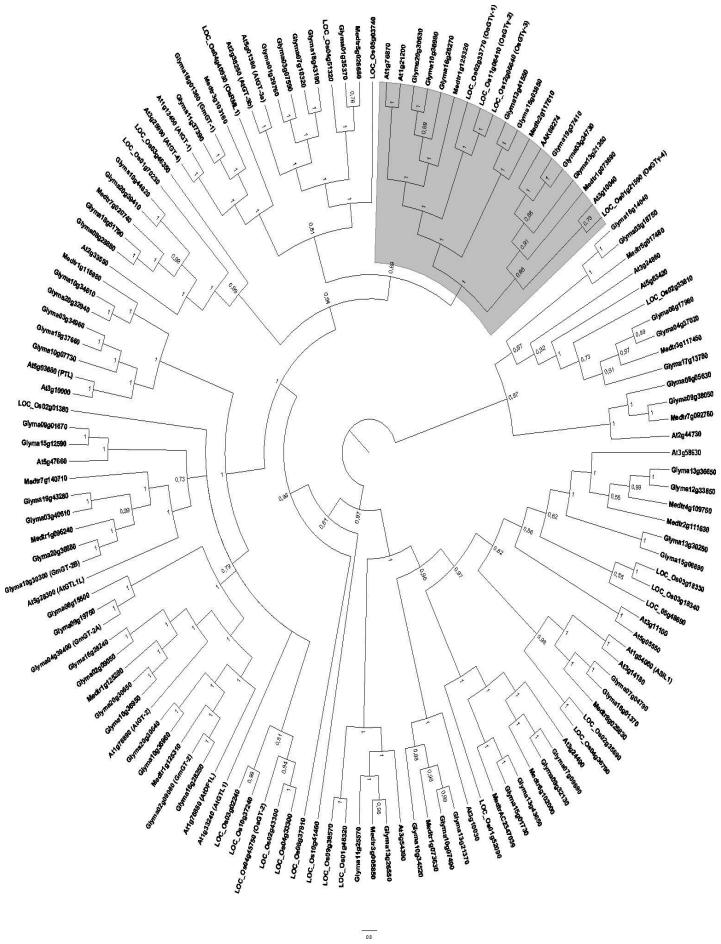

Figure 4 - Bayesian phylogenetic tree of 137 plant trihelix-GT proteins. The Bayesian analysis was conducted using Mr.Bayes v3.1.2 software after alignment of full-length trihelix-GT proteins from selected plant species using ClustalW. The unrooted cladogram was edited using Fig Tree ver. 1.3.1 software. Nodal support is given by posteriori probability values shown next to the corresponding nodes. The scale bar indicates the estimated number of amino acid substitutions per site. The gray area denotes GTγ subfamily described by Fang *et al.* (2010). Previously reported GT factors were identified according to their accession/locus numbers, the other genes were designated according to their locus ID at Phytozome. *A. thaliana* (At); *G. max* (Glyma); *Medicago truncatula* (Medtr) and *O. sativa* (LOC_Os).

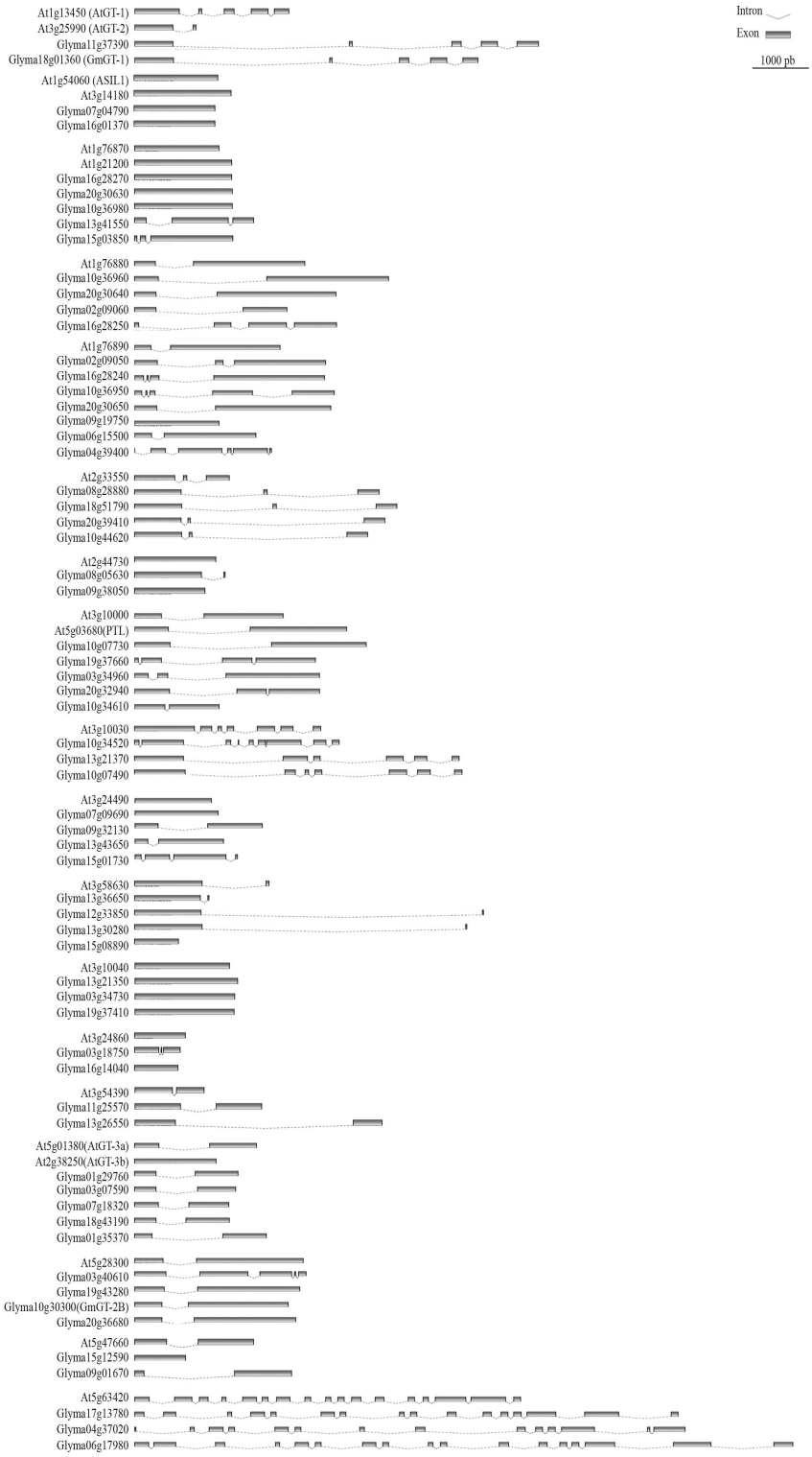

Figure 5 - Gene organization of phylogenetically related full-length coding sequences from *Arabidopsis* and soybean trihelix-GT transcription factors. Intron-exon maps were drawn using Fancy Gene ver. 1.4 software.

occurrence of an alternative splicing in Glyma19g37410 or Glyma03g34730, both considered to be phylogenetically closest to the unidentified gene locus.

Hence, when taking into account the full-length protein sequence, the GT-factor family might be divided into two subgroups, in one of these subgroups a branch corresponded to the already described subfamily γ (Figure 4, in gray). Despite the fact that subfamilies α and β were not distinguished, other probabilities supported our tree, especially when inner nodes were observed.

When gene organization among *Arabidopsis* and soybean sequences was compared (Figure 5), the number of introns ranged from zero (twenty three genes) up to 16 (At5g63420 and Glyma06g17980), and some phylogenetically close sequences showed the same gene structure. For example, the *Arabidopsis* At3g10040 and its soybean ortholog do not have intron, whereas At2g33550 and related members have two introns, with remarkable differences in intron size.

As observed for bHLH transcription factors, the soybean GT factor family encompasses almost three times more members than *Arabidopsis* or rice, a consequence of the WGD events that took place during plant evolution. In several cases, soybean paralogs clustered with one *M. truncatula* gene, indicating that these paralogs probably derived from a WGD event that occurred after the divergence of the two legume species. Similarly, Schmutz *et al.* (2010) refer to a *Glycine*-specific WGD event, estimated to have occurred about 13 million years ago. However, the possibility that extra *M. truncatula* orthologs might arise upon the completion of its genome sequencing should not be discarded.

Recently, the Os*GT*γ subfamily was proposed to participate in the regulation of stress tolerance in rice (Fang *et al.*, 2010). Os*GT*γ -1 showed more specific expression pattern than their counterparts Os*GT*γ-2 and Os*GT*γ-3, which are supposedly redundant. None of them was responsive to light, but their transcript levels increased in response to salt and cold stresses, whereas Os*GT*γ-1 was upregulated by ABA and SA stimulus. It is possible that some soybean members of this subfamily may act in response to stressor agents, but more studies are required in order to understand whether the pattern seen in rice GTγ factors also occurs in soybean and *M. truncatula.* Our analysis, so far, does not indicate their involvement in an abiotic and/or biotic stress response. Moreover, soybean genes previously reported as involved in stress responses (Xie *et al.*, 2009) together with other genes herein identified are dispersed in different tree branches, indicating that this family is in fact evolutionarily diversified.

Conclusion

The present study identified new members of soybean bHLH and trihelix-GT transcription factor families, some of which seem to be involved in responses to environmental stresses. It also emphasizes the role of duplication events in the expansion and evolution of soybean transcription factor families, indicating that exciting new layers of complexity might exist in this species' regulatory mechanisms, including biotic and abiotic stress responses.

Acknowledgments

This research project was supported by CNPq, CAPES, FAPERGS-PRONEX and GENOSOJA/CNPq.

References

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D and Shinozaki K (1997) Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9:1859-1868.

Abe H, Urao T, Ito T, Seki M and Shinozaki K (2003). *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63-78.

Anderson JP, Badruzsaufari E, Schenk PM, Manners JM and Desmond OJ (2004) Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. Plant Cell 16:3460-3479.

Amoutzias GD, Robertson DL, Oliver SG and Bornberg-Bauer E (2004) Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. EMBO Rep 5:274-279.

Brewer PB, Howles PA, Dorian K, Griffith ME, Ishida T, Kaplan-Levy RN, Kilinc A and Smyth DR (2004) PETAL LOSS, a trihelix transcription factor gene, regulates perianth architecture in the *Arabidopsis* flower. Development 131:4035-4045.

Carretero-Paulet L, Galstyan A, Roig-Villanova I, Martinez-Garcia JF, Bilbao-Castro JR and Robertson DL (2010) Genome wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss and algae. Plant Physiol 153:1398-1412.

Chandler JW, Cole M, Flier A and Werr W (2009) BIM1, a bHLH protein involved in brassinosteroid signalling, controls *Arabidopsis* embryonic patterning via interaction with DORNROSCHEN and DORNROSCHEN-LIKE. Plant Mol Biol 69:57-68.

Chen M, Xu Z, Xia L, Li L, Cheng X, Dong J, Wang Q and Ma Y (2009) Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, *GmDREB*3, in soybean (*Glycine max* L.). J Exp Bot 60:121-135.

Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, García-Casado G, López-Vidriero I, Lozano FM, Ponce MR, *et al.* (2007) The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448:666-671.

Dehesh K, Bruce WB and Quail PH (1990) A trans-acting factor that binds to a GT-motif in a phytochrome gene promoter. Science 250:1397-1399.

EMBRAPA (2007) Tecnologias de Produção de Soja. 10th edition. Embrapa Soja, Londrina, 225 pp.

Eulgem T (2005) Regulation of the *Arabidopsis* defense transcriptome. Trends Plant Sci 10:71-78.

Fang Y, Xie K, Hou X, Hu H and Xiong L (2010) Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. Mol Genet Genomics 283:157-169.

Fernández-Calvo P, Chini A, Fernández-Barbero G, Chico J-M, Gimenez-Ibanez S, Geerinck J, Eeckhout D, Schweizer F, Godoy M, Franco-Zorrilla JM, *et al.* (2011) The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell 23:701-15.

Freeling M (2009) Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol 60:433-453.

Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M and Chory J (2002) Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. Genetics 162:1445-1456.

Gao M-J, Lydiate DJ, Li X, Lui H, Gjetvaj B, Hegedus DD and Rozwadowski K (2009) Repression of seed maturation genes by a trihelix transcriptional repressor in *Arabidopsis* seedlings. Plant Cell 21:54-71.

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF and 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.

Green PJ, Kay SA and Chua N-H (1987) Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. EMBO J 6:2543-2549.

Green PJ, Yong M-H, Cuozzo M, Kano-Murakami Y, Silverstein P and Chua N-H (1988) Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the rbcS-3A gene. EMBO J 7:4035-4044.

Griffith ME, Conceição AS and Smyth DR (1999) PETAL LOSS gene regulates initiation and orientation of second whorl organs in the *Arabidopsis* flower. Development 126:5635-5644.

Haberer G, Hindemitt T, Meyers BC and Mayer KFX (2004) Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. Plant Physiol 136:3009-3022.

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B and Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol Biol Evol 20:735-747.

Hiratsuka K, Wu X, Fukuzawa H and Chua NH (1994) Molecular dissection of GT-1 from *Arabidopsis*. Plant Cell 6:1805-1813.

Hiruma K, Nishiuchi T, Kato T, Bednarek P, Okuno T, Schulze-Lefert P and Takano Y (2011) *Arabidopsis* ENHANCED DISEASE RESISTANCE 1 is required for pathogen-induced expression of plant defensins in nonhost resistance and acts through interference of MYC2-mediated repressor function. Plant J 67:980-992.

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

Kuhn RM, Caspar T, Dehesh K and Quail PH (1993) DNA binding factor GT-2 from *Arabidopsis*. Plant Mol Biol 23:337-348.

Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, *et al.* (2006) Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and *Arabidopsis*. Plant Physiol 141:1167-1184.

Li X, Qin G, Chen Z, Gu H and Qu L-J (2008) A gain-of-function mutation of transcriptional factor PTL results in curly leaves, dwarfism and male sterility by affecting auxin homeostasis. Plant Mol Biol 66:315-327.

Liu H, Yu X, Li K, Klejnot J, Yang H, Lisiero D and Lin C (2008) Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. Science 322:1535-1539.

Long TA, Tsukagoshi H, Busch W, Lahner B, Salt DE and Benfey PN (2010) The bHLH transcription factor POPEYE regulates response to iron deficiency in *Arabidopsis* roots. Plant Cell 22:2219-2236.

Lorenzo O, Chico JM and Sa JJ (2004) JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in *Arabidopsis*. Plant Cell 16:1938-1950.

Ludwig SR, Habera LF, Dellaporta SL and Wessler SR (1989) Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci USA 86:7092-7096.

Maréchal E, Hiratsuka K, Delgado J, Nairn A, Qin J, Chait BT and Chua NH (1999) Modulation of GT-1 DNA-binding activity by calcium-dependent phosphorylation. Plant Mol Biol 40:373-386.

McCarty DR and Chory J (2000) Conservation and innovation in plant signaling pathways. Cell 103:201-209.

Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, Linstead P, Schaefer DG and Dolanet L (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. Science 316:1477-1480.

Miles MR, Frederick RD and Hartman GL (2006) Evaluation of soybean germplasm for resistance to *Phakopsora pachyrhizi*. Plant Health Progr DOI 10.1094/PHP-2006-0104-01-RS.

Morgenstern B and Atchley WR (1999) Evolution of bHLH transcription factors: Modular evolution by domain shuffling? Mol Biol Evol 16:1654-1663.

Nagano Y, Inaba T, Furuhashi H and Sasaki Y (2001) Trihelix DNA-binding protein with specificities for two distinct cis-elements: Both important for light down-regulated and dark-inducible gene expression in higher plants. J Biol Chem 276:22238-22243.

Nagata T, Niyada E, Fujimoto N, Nagasaki Y, Noto K, Miyanoiri Y, Murata J, Hiratsuka K and Katahira M (2010) Solution structures of the trihelix DNA-binding domains of the wild-type and a phosphomimetic mutant of *Arabidopsis* GT-1: Mechanism for an increase in DNA-binding affinity through phosphorylation. Proteins 78:3033-3047.

Nascimento LC, Costa GGL, Binneck E, Pereira GAG and Carazzolle MF (2012) Databases and bioinformatics pipelines of the Genosoja project. Genet Mol Biol 35(suppl 1): 203-211.

Nei M and Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121-152.

O'Grady K, Goekjian VH, Naim CJ, Nagao RT and Key JL (2001) The transcript abundance of *GmGT*-2, a new member of the GT-2 family of transcription factors from soybean, is down-regulated by light in a phytochrome-dependent manner. Plant Mol Biol 47:367-378.

Park HC, Kim ML, Kang YH, Jeon JM, Yoo JH, Kim MC, Park CY, Jeong JC, Moon BC, Lee JH, *et al.* (2004) Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. Plant Physiol 135:2150-2161.

Pinheiro GL, Marques CS, Costa MDBL, Reis PAB, Alves MS, Carvalho CM, Fietto LG and Fontes EPB (2009) Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. Gene 444:10-23.

Pires N and Dolan L (2010) Origin and diversification of basic-helix-loop-helix proteins in plants. Mol Biol Evol 27:862-874.

Poppenberger B, Rozhon W, Khan M, Husar S, Adam G, Luschnig C, Fujioka S and Sieberer T (2011) CESTA, a positive regulator of brassinosteroid biosynthesis. EMBO J 30:1149-1161.

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

Rao KVM, Raghavendra AS and Reddy KJ (2006) Physiology and Molecular Biology of Stress Tolerance in Plants. Springer, Amsterdam, 345 pp.

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, *et al.* (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319:64-69.

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

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.* (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Seo J-S, Joo J, Kim M-J, Kim Y-K, Nahm BH, Song SI, Cheong J-J, Lee JS, Kim J-K and Choi YD (2011) OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. Plant J 65:907-921.

Singh KB, Foley RC and Sánchez LO (2002) Transcription factors in plant defense and stress responses. Curr Opin Plant Biol 5:430-436.

Szécsi J, Joly C, Bordji K, Varaud E, Cock JM, Dumas C and Bendahmane M (2006) BIGPETALp, a bHLH transcription factor is involved in the control of *Arabidopsis* petal size. EMBO J 25:3912-3920.

Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software ver. 4.0. Mol Biol Evol 24:1596-1599.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTALW: Improving the sensitivy of progressive multiple alignment through sequences weighting, position-specific gap penalties and wight matrix choice. Nucleic Acids Res 22:4673-4680.

Tian A-G, Wang J, Cui P, Han Y-J, Xu H, Cong L-J, Huang X-G, Wang X-L, Jiao Y-Z, Wang B-J, *et al.* (2004). Characterization of soybean genomic features by analysis of its expressed sequence tags. Theor Appl Genet 108:903-913.

Todd AT, Liu E, Polvi SL, Pammett RT and Page JE (2010) A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. Plant J 62:589-600.

Toledo-Ortiz G, Huq E and Quail PH (2003) The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell 15:1749-1770.

Wang R, Hong G and Han B (2004) Transcript abundance of rml1, encoding a putative GT1-like factor in rice, is up-regulated by *Magnaporthe grisea* and down-regulated by light. Gene 324:105-115.

Xie Z-M, Zou H-F, Lei G, Wei W, Zhou Q-Y, Niu C-F, Liao Y, Tian A-G, Ma B, Zhang W-K, *et al.* (2009). Soybean trihelix transcription factors *GmGT*-2A and *GmGT*-2B improve plant tolerance to abiotic stresses in transgenic *Arabidopsis*. PloS One 4:e6898.

Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y and Wu P (2005) OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. Plant Physiol 138:2087-2096.

Yoshioda K and Shinozaki K (2009) Signal Crosstalk in Plant Stress Responses. Wiley-Blackwell, Ames, 210 pp.

Zhang G, Chen M, Li L, Xu Z, Chen X, Guo J and Ma Y (2009) Overexpression of the soybean *GmERF*3 gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco. J Exp Bot 60:3781-3796.

Zheng L, Ying Y, Wang L, Wang F, Whelan J and Shou H (2010) Identification of a novel iron regulated basic helix-loop-helix protein involved in Fe homeostasis in *Oryza sativa*. BMC Plant Biol 10:e166.

Zhou D-X (1999) Regulatory mechanism of plant gene transcription by GT-elements and GT-factors. Trends Plant Sci 4:210-214.

Zhou Q-Y, Tian A-G, Zou H-F, Xie Z-M, Lei G, Huang J, Wang C-M, Wang H-W, Zhang J-S and Chen S-Y (2008) Soybean WRKY-type transcription factor genes, *GmWRKY*13, *GmWRKY*21, and *GmWRKY*54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants. Plant Biotechnol J 6:486-503.

Internet Resources

LGE Soybean Genome database, http://www.lge.ibi.unicamp.br/soja/ (October 21, 2011).

Fancy Gene v1.4, http://host13.bioinfo3.ifom-ieocampus.it/fancygene/ (October 21, 2011).

FigTree software, http://tree.bio.ed.ac.uk/software/figtree/ (October 21, 2011).

JGI - Joint Genome Institute, http://www.jgi.doe.gov (October 21, 2011).

LGE Soybean Genome database, http://bioinfo03.ibi.unicamp.br/soja/ (October 21, 2011).

MEME (multiple EM for motif elicitation) software, http://meme.sdsc.edu/meme4_4_0/cgi-bin/meme.cgi (October 21, 2011).

Programa Nacional de Produção e Uso de Biodiesel, http://www.biodiesel.gov.br/ (October 21, 2011).

Phytozome database, http://www.phytozome.org (October 21, 2011).

SMART database, http://smart.embl-heidelberg.de/ (October 21, 2011).

Genetics and Molecular Biology, 36, 2, 214-224 (2013)

www.sbg.org.br

Research Article

Identification of the soybean *HyPRP* family and specific gene response to Asian soybean rust disease

Lauro Bücker Neto[1], Rafael Rodrigues de Oliveira[1], Beatriz Wiebke-Strohm[1], Marta Bencke[1], Ricardo Luís Mayer Weber[1], Caroline Cabreira[1], Ricardo Vilela Abdelnoor[2], Francismar Correa Marcelino[2], Maria Helena Bodanese Zanettini[1] and Luciane Maria Pereira Passaglia[1]

[1]*Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*
[2]*EMBRAPA Soja, Londrina, PR, Brazil.*

Abstract

Soybean [*Glycine max* (L.) Merril], one of the most important crop species in the world, is very susceptible to abiotic and biotic stress. Soybean plants have developed a variety of molecular mechanisms that help them survive stressful conditions. Hybrid proline-rich proteins (HyPRPs) constitute a family of cell-wall proteins with a variable N-terminal domain and conserved C-terminal domain that is phylogenetically related to non-specific lipid transfer proteins. Members of the HyPRP family are involved in basic cellular processes and their expression and activity are modulated by environmental factors. In this study, microarray analysis and real time RT-qPCR were used to identify putative *HyPRP* genes in the soybean genome and to assess their expression in different plant tissues. Some of the genes were also analyzed by time-course real time RT-qPCR in response to infection by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust disease. Our findings indicate that the time of induction of a defense pathway is crucial in triggering the soybean resistance response to *P. pachyrhizi*. This is the first study to identify the soybean HyPRP group B family and to analyze disease-responsive *Gm*HyPRP during infection by *P. pachyrhizi*.

Keywords: fungal disease, *HyPRP* genes, *Glycine max*, real time RT-qPCR.

Received: August 20, 2012; Accepted: December 19, 2012.

Introduction

Soybean [*Glycine max* (L.) Merril], one of the most important and extensively cultivated crops in the world, is widely used for human and animal consumption because of the high protein and oil content of its seeds. Recently, soybean oil has emerged as a source of renewable fuel and its advantages over current food-based biofuels have been demonstrated (Hill *et al.*, 2006). However, unfavorable field conditions may severely restrict the soybean yield, with one of the major concerns among Brazilian soybean producers being Asian soybean rust (ASR) disease. ASR, a severe disease caused by the fungus *Phakopsora pachyrhizi*, results in significant yield losses in soybean production and is rapidly spreading around the world (Pivonia *et al.*, 2005; Carmona *et al.*, 2005).

Understanding the mechanisms that regulate the expression of stress-related genes is a fundamental issue in plant biology and is essential for the genetic improvement of soybean. As part of a study aimed at improving the ability of soybean to survive unfavorable conditions, He *et al.* (2002) analyzed the expression of a soybean gene encoding a hybrid proline-rich protein (*SbPRP*). The distribution of *SbPRP* mRNA was organ-specific and its expression was modulated by ABA (abscisic acid), circadian rhythm, salt and drought stress; there was also significant up-regulation in response to viral infection and salicylic acid.

Hybrid proline-rich proteins (HyPRPs), a subset of proline-rich proteins (PRPs), are poorly glycosylated cell wall glycoproteins specific to seed plants. HyPRPs can be classified into two groups (A and B) based on the specific position of cysteine residues in the carboxy-terminal domain that is absent in other PRP sub-classes. More specifically, group A HyPRPs have 4-6 cysteine residues whereas the group B carboxy-terminal domain has eight cysteines in a conserved pattern. The latter group of HyPRPs usually contains a signal peptide followed by a central proline-rich domain (PRD) and a hydrophobic carboxy-terminal non-repetitive domain with the eight conserved cysteine motifs, known as the eight-cysteine motif domain (8CM) (Josè-Estanyol and Puigdomènech 2000; Josè-Estanyol *et al.*, 2004; Battaglia *et al.*, 2007).

Send correspondence to Luciane Maria Pereira Passaglia. Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: luciane.passaglia@ufrgs.br.

Although huge progress has been made in understanding the molecular mechanisms underlying HyPRP action in several plants (Deutch and Winicov, 1995; Richards and Gardner, 1995; Goodwin *et al.*, 1996; Josè-Estanyol and Puigdomènech, 1998; Wilkosz and Schläppi, 2000; Bubier and Schläppi, 2004; Zhang and Schläppi, 2007; Priyanka *et al.*, 2010; Dvoráková *et al., 2011;* Huang *et al.*, 2011; Xu *et al.*, 2011), the roles of the soybean *HyPRP* gene family still remain largely unknown. The sequencing and assembly of the soybean genome (Schmutz *et al.*, 2010) may provide new approaches for identifying protein-coding *loci* possibly involved in the ability of soybean to survive stressful conditions.

In this report, we describe the identification and annotation of the soybean group B HyPRP family and its expression in different tissues based on microarray analysis. A subtractive library enriched for genes induced in response to *P. pachyrhizi* was analyzed and genes closely related to *SbPRP* were investigated in time-course real time RT-qPCR experiments in response to ASR.

Material and Methods

Annotations

In order to identify all possible soybean group B HyPRP sequences the conserved eight-cysteine motif (8CM) carboxy-terminal domain of a previously reported SbPRP (He *et al.*, 2002) was aligned (TBLASTN software) against the whole genome of Williams 82 soybean cultivar that is deposited in the Soybase and The Soybean Breeders Toolbox database. Homologous sequences with an e-value $< 1e^{-06}$ were re-aligned against the soybean genome to recover the maximum number of related proteins. All positive matches were scanned for the 8CM carboxy-terminal domain in the SMART database (with default threshold). Sequences that shared the general organization of HyPRPs were aligned by their carboxy-terminal domain in order to evaluate the presence of the eight-cysteine motif; no gaps were inserted in the conserved 8CM core. Sequences that did not fit these criteria were excluded from the analysis.

Cluster analysis

Multiple sequence alignments of the 35 soybean HyPRPs were done with the entire carboxy-terminal domain sequences (8CM) using the MUSCLE tool implemented in MEGA v.5.0 (Tamura *et al.*, 2011). Cluster analysis was done using two independent approaches: the neighbor-joining (NJ) method and the Bayesian method. The NJ method was done using MEGA v.5.0. The molecular distances of the aligned sequences were calculated according to the p-distance parameter, with gaps and missing data treated as pairwise deletions. Branch points were tested for significance by bootstrapping with 1000 replications. Bayesian analysis was done in MrBayes v.3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) with the mixed amino acid substitution model + gamma + invariant sites. Default settings were maintained, with the exception of Nchains and Nswaps that were set to eight and two, respectively. Two independent runs of 2,000,000 generations each with two Metropolis-coupled Monte Carlo Markov chains (MCMCMC) were run in parallel, each one starting from a random tree. Markov chains were sampled for every 100 generations and the first 25% of the trees were discarded as burn-in. The remaining trees were used to compute the majority rule consensus tree (MrBayes command allcompat) and the posterior probability of clades and branch lengths. The unrooted phylogenetic tree was visualized and edited using the software FigTree v.1.3.1.

Data mining

The expression profiles of the identified soybean *HyPRP* sequences that responded to infection by ASR were determined by analyzing a subtractive library. Leaves from accession PI 561356 (a resistant soybean genotype) were removed 12 to 192 h after *P. pachyrhizi* inoculation and used to construct a cDNA library. This experiment was done as part of the Genosoja project, a Brazilian soybean genome consortium, and the results can be obtained from the LGE database (http://www.lge.ibi.unicamp.br/soja/) by members of the consortium.

The gene expression patterns in six tissues (root and root tip, nodule, leaves, green pods, flower and apical meristem) were determined by microarray analysis and the results are available from Soybean Atlas hosted at the University of Missouri. Gene expression was confirmed based on EST data obtained from NCBI.

Reverse transcription and real time RT-qPCR

Soybean total RNA was extracted from leaves, closed flowers, open flowers, pods, seeds, stems and roots using TRIzol reagent (Invitrogen) and then treated with DNAse I (Promega), according to the manufacturer's specifications. The first-strand cDNA synthesis reaction was done using approximately 2 µg of DNA-free RNA, M-MLV Reverse Transcriptase system™ (Invitrogen) and a 24-oligo dT anchored primer. Real time RT-qPCR was done in a StepOne Real-time Cycler (Applied Biosystems). The PCR-cycling conditions consisted of 5 min of initial denaturation at 94 °C, 40 cycles of 10 s denaturation at 94 °C, 15 s annealing at 60 °C and 15 s extension at 72 °C, with a final extension of 2 min at 40 °C. The reaction products were identified by melting curve analysis done over the range of 55-99 °C at the end of each PCR run, with a stepwise temperature increase of 0.1 °C every s. Each reaction mixture (25 µL) contained 12.5 µL of diluted DNA template, 1 X PCR buffer (Invitrogen), 2.4 mM $MgCl_2$, 0.024 mM dNTP, 0.1 µM of each primer, 2.5 µL SYBR-Green (1:100,000; Molecular Probes Inc.) and 0.3 U of Platinum *Taq* DNA polymerase (Invitrogen). The first-strand cDNA-reaction

product (1:100) was evaluated in relative expression analyses. Technical quadruplicates were used in all real time RT-qPCR experiments and the template was omitted from negative controls. The same approach was applied to RNA extracted from soybean leaves to measure *HyPRP* expression in response to ASR.

The PCR amplification reactions were done using gene specific primers (Glyma06g07070: Forward CACCC ACTCCAACTCCATCT, Reverse GGCTTCGGAGGAG AAGGT; Glyma14g14220: Forward AAAAACTGTTCC TGCTGGCTT, Reverse TAAGGCAAACACGTGTTTA CCTAG; Glyma04g06970: Forward GTCCTCCTCCTTC TCCTCCTT, Reverse GAGCGTCACAGGTACGTTCA; Glyma17g11940: Forward GAAGGTTTGGCTGATTTG GA, Reverse AATGAACCTAACATGATGGAAGC) and the products obtained were sequenced. Sequencing was done on an ABI PRISM 3100 Genetic Analyzer automatic sequencer (Applied Biosystems) in the ACTGene Laboratory (Centro de Biotecnologia, UFRGS, RS, Brazil) using forward and reverse primers, as described by the manufacturer. Primer pairs designed to amplify an F-box and metalloprotease gene sequences were used as internal controls to normalize the amount of cDNA template present in each sample (Libault *et al.*, 2008). Relative changes in gene expression were described after comparative quantification of the target and reference gene amplified products using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The relative expression levels in soybean plants under mock or fungal infection were analyzed using Student's *t*-test with $p < 0.05$ indicating a significant difference (identified by an asterisk in the figures).

Bioassay for the analysis of *HyPRP*s expression during infection by ASR

The soybean plant reaction to ASR was evaluated by inoculating a field population of *P. pachyrhizi* spores initially collected from Brazilian soybean fields and maintained on a susceptible cultivar under greenhouse conditions until use. The experiment was done at Embrapa Soja (Londrina, PR, Brazil). Briefly, soybean plants were grown in a pot-based system and maintained in a greenhouse at 28 ± 1 °C on a 16/8 h light/dark cycle at a light intensity of 22.5 $\mu Em^{-2}/s$. The Embrapa-48 genotype was used as susceptible host as it develops a tan lesion after infection by ASR (van de Mortel *et al.*, 2007), and the PI561356 genotype was used as a resistant host in which the resistance to soybean rust is mapped on linkage group G (Abdelnoor R.V., personal communication). Uredospores were harvested from infected leaves with sporulating uredia and diluted in distilled water with 0.05% Tween-20 to a final concentration of 3×10^5 spores/mL. The spore suspension was sprayed onto three plants per pot at the V2 to V3 stage of growth. The V2 stage consists of a fully developed trifoliolate leaf at a node above the unifoliolate nodes and V3 stage is characterized by three nodes on the main stem, with fully developed leaves beginning with the unifoliolate nodes (Fehr and Caviness, 1977).

Spores were omitted in mock inoculations. After the fungal or mock inoculations, water-misted bags were placed over all plantlets for one day to aid the infection process and to prevent the cross-contamination of mock-infected plants. One trifoliolate leaf from each plant was collected at 1, 12, 24, 48, 96 and 192 h after inoculation (hai), frozen in liquid nitrogen and stored at -80 °C for RNA extraction. Three biological replicates from each genotype were analyzed for both treatments.

Results

Identification and microarray analysis of soybean HyPRP encoding genes

Annotation analysis based on the TBLASTN search of the 8CM carboxy-terminal domain of a previously reported *SbPRP* against Williams 82 soybean cultivar coding sequences in the Soybase and The Soybean Breeders Toolbox database identified 35 GmHyPRP-encoding genes in the soybean genome. The *GmHyPRP* genes were located in ten chromosomes, with protein sequences ranging in size from 120 to 385 amino acids. Chromosome 17 contained the highest number of *GmHyPRP* genes (10 out of 35), whereas only a single gene was detected in each of chromosomes 1, 4, 6 and 14. Figure 1 shows the relative locations of the genes on their respective chromosomes and genes located at *loci* close to each other are indicated as possible tandem duplications. A standardized nomenclature based on the gene order in the chromosomes was used for all *GmHyPRP* genes identified in this work. This same approach has recently been used by other researchers to facilitate the description of their findings (Table 1).

The previously reported *SbPRP* gene corresponds to the gene model Glyma14g14220 in the Williams 82 genome and, based on our criteria, was identified as *GmHyPRP16*. Only two gene models, corresponding to Glyma20g06290 (*GmHyPRP33*) and Glyma20g35080 (*GmHyPRP35*), were corrected manually and, based on the genomic sequence, one of them (Glyma20g35080) showed two possible open reading frames (ORFs), with or without the presence of an intron. However, a gene model without introns became more probable when all *HyPRP* cDNA sequence encoding proteins were analyzed, since none of the corresponding genes contained introns in their genomic sequences. Among the annotated genes, 29 had corresponding expressed sequence tags (ESTs) and 27 had their full length proteins confirmed, indicating that they are unlikely to be pseudogenes. Only for six genes were there no ESTs in either of the databases analyzed.

All soybean HyPRPs had an N-terminal secretion signal, except for *GmHyPRP34* in which the peptide signal was replaced by a low complexity region. Since this protein was more related to a HyPRP than to any other class of cell

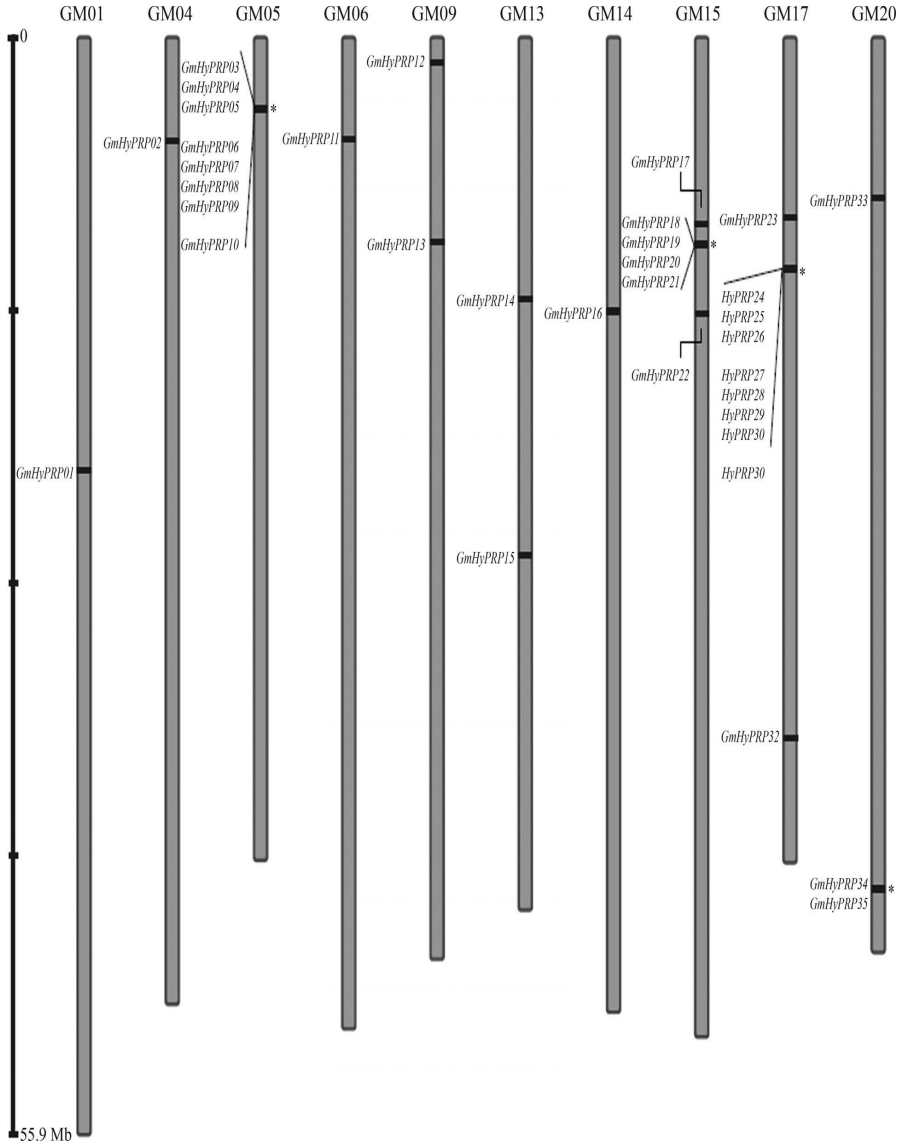

Figure 1 - Representation of the locations for *GmHyPRP* genes on each soybean chromosome. The asterisks indicate possible tandem duplicated genes. Gm indicates chromosome numbers.

wall proteins (data not shown), in the present study the corresponding gene was considered to be a member of the soybean *HyPRP* gene family. The sequences for *GmHyPRP08*, *GmHyPRP14*, *GmHyPRP15*, *GmHyPRP29*, *GmHyPRP23* and *GmHyPRP33* belong to the conserved-type (C-type) HyPRPs and those for *GmHyPRP04* and *GmHyPRP25* contain glycine-rich N-terminal domains. In the first group, the 8CM cluster analysis formed a stable branch in the tree, but this was not the case for the second group (Figure 2, left side; Supplementary Material Figure S1).

Expression of the soybean *GmHyPRP* gene family was initially analyzed in response to ASR disease by mining a subtractive library in order to identify responsive genes. Six genes were up-regulated during infection by *P. pachyrhizi* (Figure 2, middle). *GmHyPRP15* and *GmHyPRP29* coded for soybean C-type HyPRPs while the other four genes (*GmHyPRP02*, *GmHyPRP11*, *GmHyPRP16* and *GmHyPRP32*) formed a stable branch in which all members responded to the pathogen.

The expression profile of the 35 soybean genes identified as described above was assessed in six vegetative plant organs: root and root tip, nodule, leaves, green pods, flower and apical meristem (Figure 2, right side). Three genes (*GmHyPRP22*, *GmHyPRP34* and *GmHyPRP35*) were not

Table 1 - Annotation of soybean HyPRP-encoding genes. Gene nomenclature was based on chromosomal order[1].

Accession number in Phytozome (gene)	Proposed name	Chromosome	CDS/ORF (bp)	Expression confirmed by EST (GenBank accession number)	Full-length protein confirmed by cDNA
Glyma01g17820	*GmHyPRP01*	1	387	BQ273195.1	+
Glyma04g06970	*GmHyPRP02*	4	534	EV274219.1	+
Glyma05g04380	*GmHyPRP03*	5	414	EV263905.1	+
Glyma05g04390	*GmHyPRP04*	5	519	AI496419.1	+
				BF595475.1	
Glyma05g04400	*GmHyPRP05*	5	411	EV278968.1	+
Glyma05g04430	*GmHyPRP06*	5	405	CA784637.1	+
Glyma05g04440	*GmHyPRP07*	5	411	EV271119.1	+
Glyma05g04450	*GmHyPRP08*	5	540	AW569247.1	-
Glyma05g04460	*GmHyPRP09*	5	381	-	-
Glyma05g04490	*GmHyPRP10*	5	396	BG511695.1	+
Glyma06g07070	*GmHyPRP11*	6	666	BI945945.1	+
				AW279308.1	
Glyma09g01680	*GmHyPRP12*	9	387	FK021328.1	+
Glyma09g10340	*GmHyPRP13*	9	375	FK001188.1	+
Glyma13g11090	*GmHyPRP14*	13	1155	AW152930.1	+
				GR835813.1	
				BG649969.1	
Glyma13g22940	*GmHyPRP15*	13	684	EV278617.1	+
Glyma14g14220[2]	*GmHyPRP16*	14	381	EV274235.1	+
Glyma15g12600	*GmHyPRP17*	15	384	AW278280.1	+
Glyma15g13740	*GmHyPRP18*	15	360	-	-
Glyma15g13750	*GmHyPRP19*	15	360	AW277674.1	+
Glyma15g13760	*GmHyPRP20*	15	387	-	-
Glyma15g13770	*GmHyPRP21*	15	390	AW156395.1	-
Glyma15g17570	*GmHyPRP22*	15	420	-	-
Glyma17g11940	*GmHyPRP23*	17	573	EV280964.1	+
Glyma17g14840	*GmHyPRP24*	17	408	FK018257.1	+
Glyma17g14850	*GmHyPRP25*	17	513	FK014996.1	+
Glyma17g14860	*GmHyPRP26*	17	411	BQ453492.1	+
Glyma17g14880	*GmHyPRP27*	17	417	BU083296.1	+
Glyma17g14890	*GmHyPRP28*	17	414	BE347345.1	+
Glyma17g14900	*GmHyPRP29*	17	537	AW398015.1	+
Glyma17g14910	*GmHyPRP30*	17	396	EV268166.1	+
Glyma17g14930	*GmHyPRP31*	17	396	EV271098.1	+
Glyma17g32100	*GmHyPRP32*	17	381	BE347495.1	+
Glyma20g06290[3]	*GmHyPRP33*	20	987	BM886103.1	+
				BF070112.1	
Glyma20g35070	*GmHyPRP34*	20	369	-	-
Glyma20g35080[3 4]	*GmHyPRP35*	20	408/360	-	-

Soybean *HyPRP*-encoding gene annotation was based on Phytozome gene models. The expression data were obtained from the NCBI database.
[1]The same approach was recently used by Le *et al.* (2011).
[2]Previously reported as *ShPRP* (soybean proline-rich protein) by He *et al.* (2002).
[3]Indicates a correction in the Phytozome gene models.
[4]Based on the gene sequence Glyma20g35080 has two possible ORFs (with or without introns).

detected in any tissue. The other genes exhibited variable expression patterns. For example, *GmHyPRP06*, *GmHyPRP08*, *GmHyPRP09*, *GmHyPRP20* and *GmHyPRP27* were expressed in specific organs with differing transcript levels. A low, *ubiquitous* expression was observed for *GmHyPRP30* while the opposite was true for *GmHyPRP15*, *GmHyPRP23* and *GmHyPRP14* (C-type), all of which exhibited a high, *ubiquitous* expression in *all*

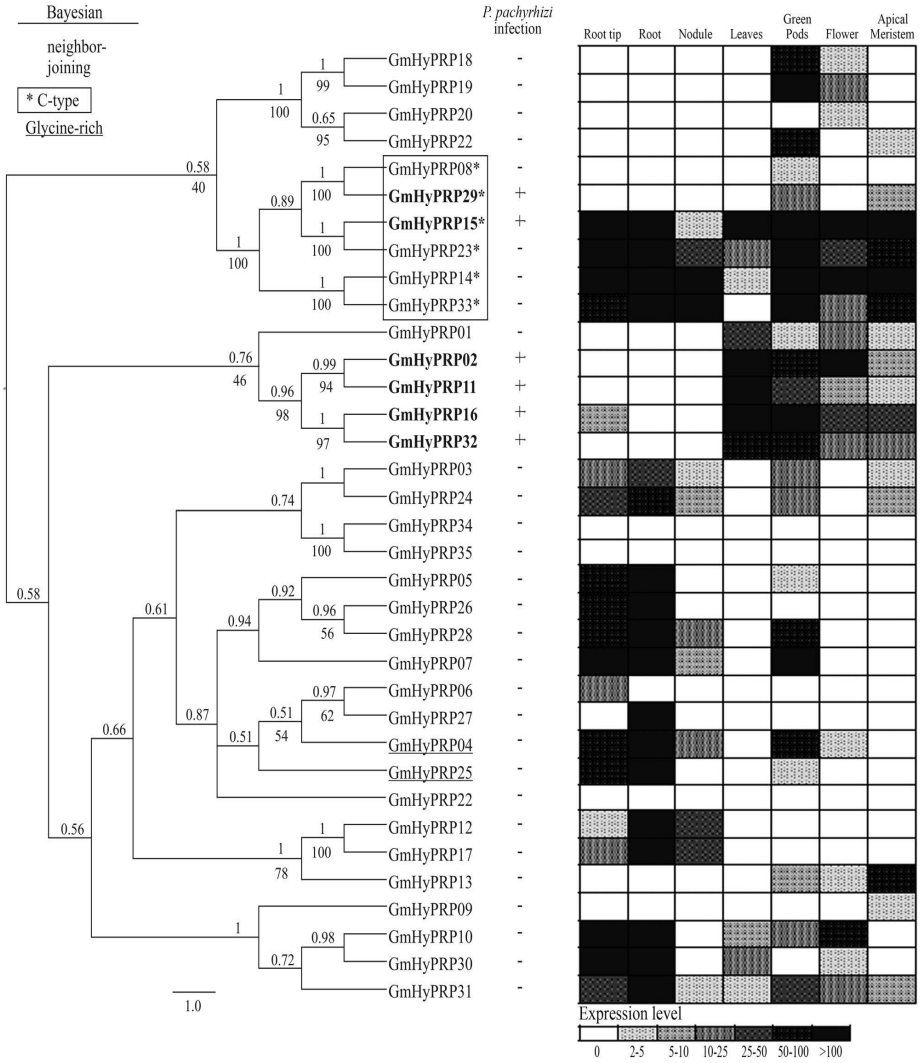

Figure 2 - Cluster analysis and expression patterns of soybean HyPRPs. Left - Bayesian cladogram of 35 soybean HyPRP proteins. The Bayesian analysis was done using Mr. Bayes v.3.1.2, after alignment of the conserved C-terminal domains of HyPRPs using Muscle. The unrooted cladogram was edited using FigTree v.1.3.1. Nodal support is given by the posteriori probability values above the branches. Numbers below the branches denote bootstrap values obtained for the same input data using neighbor-joining analysis in MEGA. The scale bar indicates the estimated number of amino acid substitutions per site. The genes were designated according to their locus ID in Phytozome. C-type proteins are shown in blue, glycine-rich N-terminal domains in red and genes responsive to ASR in bold. Middle - HyPRP expression [absence (-); presence (+)] in leaves from PI561356 (resistant genotype) infected with *P. pachyrhizi* (12-192 h). The data were obtained from subtractive library experiments available at www.lge.ibi.unicamp.br/soja/. Right - Microarray analysis of the expression profiles in root, root tip, nodule, leaves, green pods, flower and apical meristem of soybean plants. Data available at http://digbio.missouri.edu/soybean_atlas/.

organs examined. The genes in the branch responsive to infection by *P. pachyrhizi* (*GmHyPRP02*, *GmHyPRP11*, *GmHyPRP16* and *GmHyPRP32*) were almost exclusively highly expressed in leaves; *GmHyPRP29* was not expressed in leaves whereas *GmHyPRP15* had a more ubiquitous expression.

To confirm the array results for *GmHyPRP16* and its paralogs, gene expression was measured by real time RT-qPCR in different soybean tissues (Figure 3). The four genes screened were detected in almost all tissues tested. *GmHyPRP11* had a tissue-specific expression pattern and was not detected in flowers (either opened or closed).

Time-course of HyPRP gene response to infection by *P. pachyrhizi*

Since *GmHyPRP16* and its paralogs were responsive in an ASR subtractive library and since all of them were expressed in leaves, real time RT-qPCR was used to analyze their transcript levels in soybean plants inoculated with *P. pachyrhizi*. A time-course experiment was used to examine the *GmHyPRP02*, *GmHyPRP11*, *GmHyPRP16* and *GmHyPRP32* expression pattern in leaves of the highly susceptible soybean genotype Embrapa-48 and in the more disease-resistant genotype PI561356 (Figure 4). In view of the difficulty in detecting *GmHyPRP11* cDNA, this gene was analyzed at only two time points. Figure 4 shows that the susceptible soybean host *HyPRP* transcripts were significantly up-regulated at 24 h post-infection, with an additional increase, especially in *SbPRP GmHyPRP16*, at 192 h post-infection. In contrast, in the resistant soybean host, the expression of *HyPRP* transcripts was already strongly up-regulated 12 h after fungus inoculation and in all cases anticipated the gene response to infection by *P. pachyrhizi*. These plants exhibited less induction when compared to a susceptible genotype, with higher fold change occurring in *GmHyPRP32* (192 h post-infection). The response to ASR also involved the expression of *GmPR*4 (Glyma19g43460) (data not shown).

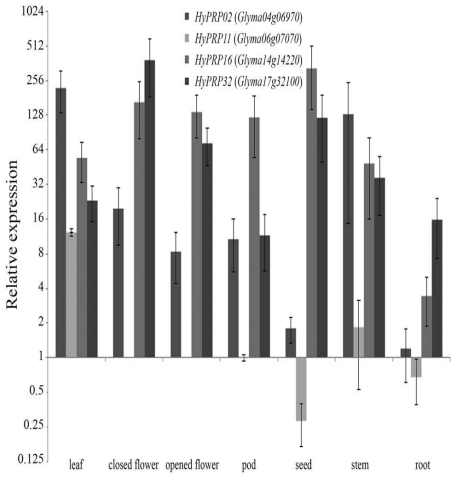

Figure 3 - Expression profile of four soybean *HyPRP*-encoding genes in different plant tissues as assessed by real time RT-qPCR. The level of expression is shown relative to that of Glyma06g07070 in pods. The columns are the mean of three biological samples (pool of three plants each sample). Y bar indicates the standard error of the mean.

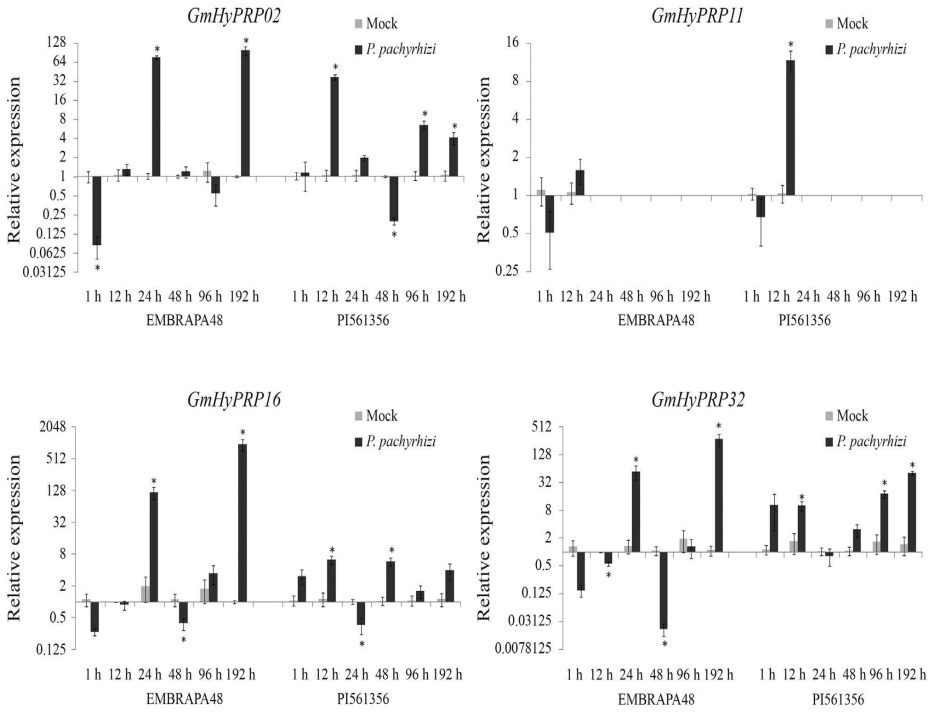

Figure 4 - Expression profile of four soybean *HyPRP*-encoding genes in response to infection by *Phakopsora pachyrhizi* in the highly susceptible genotype Embrapa-48 and in the resistant genotype PI561356. Expression was assessed by real time RT-qPCR and is shown relative to the levels of F-box and metalloprotease. The columns are the mean of three biological samples (pool of three plants each sample). Y bar indicates the standard error of the mean. Asterisk (*) indicates $p < 0.05$ compared to mock.

Discussion

HyPRP organization and expression pattern

Soybean is a palaeotetraploid genome with two major duplication events dated to about 44 and 15 million years ago (Schlueter *et al.*, 2004). Soybean was the first legume species sequenced (Schmutz *et al.*, 2010) and its genome contains 950 megabases distributed in 20 chromosomes and > 46,000 protein-coding genes. During evolution polyploidy has had a deep effect on the soybean genome structure and organization and has contributed to the emergence of duplicated gene blocks that have been retained and remain active (Schmutz *et al.*, 2010). Previous studies indicated that the genus *Glycine* has approximately twice as many chromosomes as its relatives (Doyle *et al.*, 2004). Large scale analysis has shown that ~75% of soybean genes are present in multiple copies. Diversification and gene loss, as well as chromosomal rearrangements, have modified the genomic structure over time (Schmutz *et al.*, 2010). Zhu *et al.* (1994) estimated that 25% of duplicated genes have been lost since the last polyploidization event. EST analysis indicated that each soybean gene family consists of on average 3.1 members, a smaller number than would be expected if all copies from two duplication events were retained and expressed (Nelson and Shoemaker, 2006). However, the survival rates of duplicated gene classes vary, with some being more prone to retention than others. Gene families are retained and tend to grow if they have structural and/or functional features that allow diverse functions or undergo rapid subfunctionalization (Adams and Wendel, 2005; Lan *et al.*, 2009).

To gain insight into the evolutionary dynamics of the soybean HyPRP family a phylogenetic analysis of their corresponding amino acid sequences was done using the entire carboxy-terminal domain (8CM) from *Cucumis sativus* (cucumber), *Glycine max, Medicago truncatula* and *Prunus persica* (peach) (Figure S2). Analysis of the 81 genes recovered from the databank revealed that soybean had the highest number of members, indicating that genome duplication events probably contributed to a greater number of genes than in the other species analyzed here.

We identified 35 soybean HyPRP-encoding genes that are widely distributed among plant chromosomes (1, 4, 5, 6, 9, 13, 14, 15, 17 and 20) and are arranged in tandem on chromosomes 5, 15, 17 and 20. This structural organization is characteristic of several cell wall glycoprotein-encoding genes in other species, such as *Arabidopsis thaliana* and *Oryza sativa* (rice) (Jose-Estanyol *et al.*, 2004; Sampedro *et al.*, 2005). HyPRP families with multiple copies have been described in other species (Dvorakova *et al.*, 2007) and the large number of genes found in soybean agrees with the number expected for cell wall glycoproteins in plants, *e.g.*, expansin-like A protein, that has 26 members in *A. thaliana* and 34 members in *O. sativa* (Sampedro *et al.*, 2005).

Possibly the most striking feature of the 35 soybean *HyPRPs* was the complete absence of introns in their genetic structure. Jain *et al.* (2011) have demonstrated that intronless genes constitute a significant portion of the rice (19.9%) and *Arabidopsis* (21.7%) genomes and are associated with different cellular roles and gene ontology categories. Rapidly regulated genes may have lower intron densities and is crucial for rapid gene regulation during stress, cell proliferation, differentiation, or even during development. In this context, introns can delay appropriate regulatory responses, which may explain their absence from these sequences (Jeffares *et al.*, 2008). Since HyPRPs are involved in a broad spectrum of plant responses to abiotic, biotic and developmental processes it is not surprising that a rapid adjustment in gene expression could help to overcome environmental challenges.

The N-terminal domain of known HyPRPs is highly variable in size and amino acid composition, probably because its repetitive nature allows it to undergo rearrangement (Fischer *et al.*, 2002). In such cases, phylogenetic analyses based on a single domain rather than the full-length protein appear to be more reliable, despite the domains small size and poor sequence conservation (Brinkman and Leipe, 2001). As described here, the 8CM motif was examined to establish a relationship between soybean HyPRPs and their counterparts in other plants. This domain is widely distributed in seed plants and is shared by 2S-albumins, lipid transfer proteins (LTP), HyGRPs (hybrid glycine-rich proteins), amylase and trypsin inhibitors, and group B HyPRPs. The 8CM domain is involved in a variety of functions such as seed storage, enzymatic protection and inhibition, lipid transfer and cell wall structure (José-Estanyol *et al.*, 2004). Since protein groups with distinct functions show high structural similarity with the 8CM domain it has been proposed that they share a common ancestral gene that accumulated modifications without altering the basic protein organization and acquired new functions over time (Henrissat *et al.*, 1988). During plant evolution, the first HyPRP was possibly derived from an LTP that incorporated a proline-rich N-terminal domain by gene fusion or by the introduction of a repetitive element that became shorter and that was occasionally replaced by the glycine-rich domain (Dvorakova *et al.*, 2007). Evolutionary history explains how sequences with N-terminal domains rich in glycine (*GmHyPRP04* and *GmHyPRP25*) form a stable relationship with typical HyPRPs since unconventional N-terminal domains appear to occur in a repetitive and independent manner, indicating their polyphyletic origin (as shown by cluster analysis). Even a sequence without a signal peptide (*GmHyPRP34*) proved to be closer to HyPRPs than to other related proteins. This has never been described before and could be an artifact since the respective gene was not detected in the expression database, *i.e.*, it could be a pseudogene.

C-type HyPRP proteins are a specific group of proteins with an N-terminal that is unusual in length and has a high content of hydrophobic residues. Soybean proteins that share these characteristics form a stable branch, as shown by cluster analysis. Even when the respective genes were analyzed together with those of other species they remained in the same branch (Figure S2). These proteins may be less divergent because they are ubiquitously expressed (Dvorakova *et al.*, 2007), as was the case for *GmHyPRP14*, *GmHyPRP15*, *GmHyPRP23* and *GmHyPRP33* in this study. On the other hand, microarray experiments indicated that *HyPRP08* and *HyPRP29* had a distinct expression pattern. Interestingly, both of these proteins had the smallest N-terminal domain among soybean C-type HyPRPs (data not shown).

The overall gene expression in several soybean tissues (Figure 2 - right side, and Figure 3) revealed that in some cases duplicated members had overlapping specificities and similar activities. Other related paralogs diverged in their gene expression patterns. Modifications in the *cis*-regulatory elements of promoter regions could lead to transcriptional neofunctionalization or subfunctionalization (Haberer *et al.*, 2004), which in turn could explain the similar or divergent responses in different plant tissues or even in response to the same stressor stimulus, *e.g.*, *HyPRP* genes that maintain promoter recognition sites related to plant defense (GT1GMSCAM4 and WBOXATNPR1 identified upstream of the start of transcription; data not shown) and that are responsive to infection by *P. pachyrhizi*. Further studies involving promoter transformation to verify inducible expression patterns may clarify the involvement of duplicated genes in stress-related responses.

Response of soybean cultivars to infection by *P. pachyrhizi*

Phakopsora pachyrhizi induces biphasic global gene expression in response to ASR disease. The first peak of gene expression occurs during early infection and is a non-specific defense response similar to pathogen triggered immunity (PTI). The second peak of gene expression coincides with haustoria formation and effector secretion and is consistent with the activation of *RPP2*- and *RPP3*-mediated resistance (Mortel *et al.*, 2007; Panthee *et al.*, 2007; Schneider *et al.*, 2011).

Twelve hours after fungal infection, when the early processes of apressorium formation and epidermal cell penetration occurred, the tolerant soybean genotype (PI561356) presented an up-regulation in *HyPRP* transcript levels whereas in the susceptible cultivar (Embrapa-48) no similar change was detected. The Embrapa-48 response occurred only 24 h after pathogen inoculation. Since the soybean HyPRP-encoding genes analyzed showed an expression peak in the first hours after fungal infection, we postulate that they might be involved in a non-specific defense response. The intense but late *HyPRP* expression in Embrapa-48 cultivar could be a decisive factor involved in plant susceptibility to pathogen attack since experiments based on global expression analysis suggest that the timing and the degree of induction of a defense pathway are pivotal in inducing the soybean resistance response to *P. pachyrhizi* (Mortel *et al.*, 2007; Choi *et al.*, 2008; Goellner *et al.*, 2010; Schneider *et al.*, 2011). A delayed attempt to block fungal invasion may not be as effective in stopping the infection as a less intense but early gene upregulation, such as observed in the resistant PI561356 genotype. Gene expression is reportedly faster and of greater magnitude in the incompatible interaction (Mortel *et al.*, 2007; Panthee *et al.*, 2007; Schneider *et al.*, 2011).

Some cell wall proteins, *e.g.*, extensins and proline-rich proteins (PRP), can respond promptly to pathogens, probably by enhancing physical barriers (Showalter, 1993; Schnabelrauch *et al.*, 1996). The extensins are hydroxyproline-rich glycoproteins (HRGPs) involved in cell wall self-organization during stress (Cannon *et al.*, 2008) and it seems reasonable to suggest that GmHyPRPs may have an equivalent function through modification of the cell wall structure during ASR infection. HyPRPs were recently shown to be associated with cell-wall extension processes (Dvoráková *et al.*, 2011). A subcellular localization experiment also indicated that at least *HyPRP16* was secreted into the cell wall (Figure S3) where it possibly contributed to a defense mechanism against pathogen attack, perhaps by providing more than just a mechanical barrier.

Soria-Guerra *et al.* (2010) reported that HRGP transcript levels were upregulated in susceptible and resistant genotypes of *Glycine tomentella* during infection by *P. pachyrhizi*. Microarray experiments have demonstrated that several cell wall genes among those that encode for PRPs and HRGPs were upregulated in response to nematode invasion of the soybean root system (Khan *et al.*, 2004). Even a role as one component in the defense signaling cascade cannot be ruled out since *A. thaliana* AZI1 (a HyPRP) has been shown to be involved in plant defense to ASR (Jung *et al.*, 2009).

This work is the first to identify the soybean HyPRP group B family and to analyze disease-responsive GmHyPRP during infection by *P. pachyrhizi*. Our results indicate that the time of induction of a defense pathway is crucial to triggering the soybean resistance response to *P. pachyrhizi*, the causal agent of ASR. Future studies will improve our understanding of the relationship between the proteins described here and their role(s) in adaptation to biotic stress. Such information will provide a valuable genetic resource for engineering tolerance in soybean crops.

Acknowledgments

This research was supported by grants from the Brazilian Soybean Genome Consortium (Genosoja Project), Conselho Nacional de Desenvolvimento Científico e Tec-

nológico (CNPQ) and BIOTECSUR. We thank Henrique Beck Biehl of the Centro de Microscopia Eletrônica (UFRGS) for his help with the confocal microscopy analysis and Silvia Nair Cordeiro Richter for her help with the picture editing.

References

Adams KL and Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Biol 8:135-141.

Battaglia M, Solorzano RM, Hernandez M, Cuellar-Ortiz S, Garcia-Gomez B, Marquez J and Covarrubias AA (2007) Proline-rich cell wall proteins accumulate in growing regions and phloem tissue in response to water deficit in common bean seedlings. Planta 225:1121-1133.

Brinkman FSL and Leipe DD (2001) Phylogenetic analysis. In: Baxevanis AD and Ouelette BFF (eds) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. Wiley, New York, pp 323-358.

Bubier J and Schläppi M (2004) Cold induction of EARLI1, a putative *Arabidopsis* lipid transfer protein, is light and calcium dependent. Plant Cell Environ 27:929-936.

Cannon MC, Terneus K, Hall Q, Tan L, Wang Y, Wegenhart BL, Chen L, Lamport DTA, Chen Y and Kieliszewski MJ (2008) Self-assembly of the plant cell wall requires an extensin scaffold. Proc Natl Acad Sci USA 105:2226-2231.

Carmona MA, Gally ME and Lopez SE (2005) Asian soybean rust: Incidence, severity, and morphological characterization of *Phakopsora pachyrhizi* (Uredinia and Telia) in Argentina. Plant Dis 89:109.

Choi JJ, Alkharouf NW, Schneider KT, Matthews BF and Frederick RD (2008) Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. Funct Integr Genomic 8:341-359.

Deutch CE and Winicov I (1995) Post-transcriptional regulation of a salt-inducible alfalfa gene encoding a putative chimeric proline-rich cell wall protein. Plant Mol Biol 27:411-418.

Doyle JJ, Doyle JL, Rauscher JT and Brown AHD (2004) Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): A study of contrasts. Biol J Linn Soc 82:583-597.

Dvoráková L, Cvrckova F and Fischer L (2007) Analysis of the hybrid proline-rich protein families from seven plant species suggests rapid diversification of their sequences and expression patterns. BMC Genomics 8:e412.

Dvoráková L, Srba M, Opatrny Z and Fischer L (2011) Hybrid proline-rich proteins: Novel players in plant cell elongation? Ann Bot 109:453-462.

Fehr WR and Caviness CE (1977) Stages on Soybean Development. Spec. Rep. No. 80. Coop. Ext. Serv., Agric. and Home Econ. Expn. Stn., Iowa State Univ., Ames, 21 pp.

Fischer L, Lovas Á, Opatrny Z and Bánfalvi Z (2002) Structure and expression of a hybrid proline-rich protein gene in the solanaceous species, *Solanum brevidens*, *Solanum tuberosum*, and *Lycopersicum esculentum*. J Plant Physiol 159:1271-1275.

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

Goodwin W, Pallas JA and Jenkins GI (1996) Transcripts of a gene encoding a putative cell wall-plasma membrane linker protein are specifically cold-induced in *Brassica napus*. Plant Mol Biol 31:771-781.

Haberer G, Hindemitt T, Meyers BC and Mayer KF (2004) Transcriptional similarities, dissimilarities, and conservation of *cis*-elements in duplicated genes of *Arabidopsis*. Plant Physiol 136:3009-3022.

He CY, Zhang JS and Chen SY (2002) A soybean gene encoding a proline-rich protein is regulated by salicylic acid, an endogenous circadian rhythm and by various stresses. Theor Appl Genet 104:1125-1131.

Henrissat B, Popineau Y and Kader JC (1988) Hydrophobic-cluster analysis of plant protein sequences. A domain homology between storage and lipid-transfer proteins. Biochem J 255:901-905.

Hill J, Nelson E, Tilman D, Polasky S and Tiffany D (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci USA 103:11206-11210.

Huang G, Gong S, Xu W, Li P, Zhang D, Qin L, Li W and Li X (2011) GhHyPRP4, a cotton gene encoding putative hybrid proline-rich protein, is preferentially expressed in leaves and involved in plant response to cold stress. Acta Biochim Biophys Sin (Shanghai) 43:519-527.

Huelsenbeck JP and Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Jain M, Khurana P, Tyagi AK and Khurana JP (2011) Genome-wide analysis of intronless genes in rice and *Arabidopsis*. Funct Integr Genomic 8:69-78.

Jeffares DC, Penkett CJ and Bähler J (2008) Rapidly regulated genes are intron poor. Trends Genet 24:375-378

Josè-Estanyol M and Puigdomènech P (1998) Developmental and hormonal regulation of genes coding for proline-rich proteins in female inflorescences and kernels of maize. Plant Physiol 116:485-494.

Josè-Estanyol M and Puigdomènech P (2000) Plant cell wall glycoproteins and their genes. Plant Physiol Biochem 38:97-108.

Josè-Estanyol M, Gomis-Ruth FX and Puigdomenech P (2004) The eight-cysteine motif, a versatile structure in plant proteins. Plant Physiol Biochem 42:355-365.

Josè-Estanyol M, Perez P and Puigdomenech P (2005) Expression of the promoter of HyPRP, an embryo-specific gene from *Zea mays* in maize and tobacco transgenic plants. Gene 356:146-152.

Jung HW, Tschaplinski TJ, Wang L, Glazebrook J and Greenberg JT (2009) Priming in systemic plant immunity. Science 324:89-91.

Khan R, Alkharouf N, Beard H, Macdonald M, Chouikha I, Meyer S, Grefenstette J, Knap H and Matthews B (2004) Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. J Nematol 36:241-248.

Lan T, Yang Z-L, Yang X, Liu Y-J, Wang X-R and Zeng Q-Y (2009) Extensive functional diversification of the *Populus* glutathione S-transferase supergene family. Plant Cell Online 21:3749-3766.

Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K and Tran LS (2011) Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Res 4:263-276.

Libault M, Thibivilliers S, Bilgin DD, Radwan O, Benitez M, Clough SJ and Stacey G (2008) Identification of four soybean reference genes for gene expression normalization. Plant Gen 1:44-54.

Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402-408.

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

Nelson RT and Shoemaker R (2006) Identification and analysis of gene families from the duplicated genome of soybean using EST sequences. BMC Genomics 7:204.

Panthee DR, Yuan JS, Wright DL, Marois JJ, Mailhot D and Stewart Jr CN (2007) Gene expression analysis in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) in an early growth stage. Funct Integr Genomic 7:291-301.

Pivonia S, Yang XB and Pan Z (2005) Assessment of epidemic potential of soybean rust in the United States. Plant Dis 89:678-682.

Priyanka B, Sekhar K, Reddy VD and Rao KV (2010) Expression of pigeonpea hybrid-proline-rich protein encoding gene (CcHyPRP) in yeast and *Arabidopsis* affords multiple abiotic stress tolerance. Plant Biotechnol J 8:76-87.

Richards KD and Gardner RC (1995) pEARLI1 (Accession No. L43080): An *Arabidopsis* member of a conserved gene family (PGR95-099). Plant Physiol 109:1497.

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

Sampedro J, Lee Y, Carey RE, DePamphilis C and Cosgrove DJ (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. Plant J 44:409-419.

Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ and Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome 47:868-876.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.* (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Schnabelrauch LS, Kieliszewski M, Upham BL, Alizedeh H and Lamport DT (1996) Isolation of pl 4.6 extensin peroxidase from tomato cell suspension cultures and identification of Val-Tyr-Lys as putative intermolecular cross-link site. Plant J 9:477-489.

Schneider KT, van de Mortel M, Bancroft TJ, Braun E, Nettleton D, Nelson RT, Frederick RD, Baum TJ, Graham MA, Whitham SA (2011) Biphasic gene expression changes elicited by *Phakopsora pachyrhizi* in soybean correlate with fungal penetration and haustoria formation. Plant Physiol 157:355-371.

Showalter AM (1993) Structure and function of plant cell wall proteins. Plant Cell 5:9-23.

Soria-Guerra RE, Rosales-Mendoza S, Chang S, Haudenshield JS, Padmanaban A, Rodrigues-Zas S, Hartman GL, Ghabrial S and Korban SS (2010) Transcriptome analysis of resistant and susceptible genotypes of *Glycine tomentella* during *Phakopsora pachyrhizi* infection reveals novel rust resistance genes. Theor Appl Genet 40:1432-2242.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739.

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

Wilkosz R and Schläppi M (2000) A gene expression screen identifies EARLI1 as a novel vernalization-responsive gene in *Arabidopsis thaliana*. Plant Mol Biol 44:777-787.

Xu D, Huang X, Xu ZQ and Schläppi M (2011) The HyPRP gene EARLI1 has an auxiliary role for germinability and early seedling development under low temperature and salt stress conditions in *Arabidopsis thaliana*. Planta 234:565-577.

Zhang Y and Schläppi M (2007) Cold responsive EARLI1 type HyPRPs improve freezing survival of yeast cells and form higher order complexes in plants. Planta 227:233-243.

Zhu T, Schupp JM, Oliphant A and Keim P (1994) Hypomethylated sequences: Characterization of the duplicated soybean genome. Mol Gen Genet 244:638-645.

Internet Resources

Soybase and The Soybean Breeders Toolbox database, http://soybase.org/ (accessed in July 6, 2011).

FigTree v.1.3.1, http://tree.bio.ed.ac.uk/software/figtree/ (accessed in July 6, 2011).

Genosoja project, LGE database, http://www.lge.ibi.unicamp.br/soja/ (accessed in July 6, 2011).

Soybean gene *expression patterns* in tissues in Soybean Atlas, http://digbio.missouri.edu/soybean_atlas/ (accessed in July 6, 2011).

Supplementary Material

The following online material is available for this article:

- Figure S1 - Alignment of the conserved C-terminal domains of soybean *HyPRPs* using Muscle software.
- Figure S2 - Bayesian phylogenetic tree of 81 *HyPRPs* from soybean and three other plant species.
- Figure S3 - Subcellular localization of GmHyPRP16 in soybean root cells after dehydration.

This material is available as part of the online article from http://www.scielo.br/gmb.

Associate Editor: Everaldo Gonçalves de Barros