PRISCILA ALVES DA SILVA

VETOR DE SILENCIAMENTO GÊNICO E ANÁLISE COMPREENSIVA DA RESPOSTA AO ESTRESSE NO RETÍCULO ENDOPLASMÁTICO EM SOJA

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

VIÇOSA
MINAS GERAIS – BRASIL
2015

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
S586v
2015

Silva, Priscila Alves da, 1981-
Vetor de silenciamento gênico e análise compreensiva da resposta ao estresse no retículo endoplasmático em soja / Priscila Alves da Silva. – Viçosa, MG, 2015.
x, 140f. : il. (algumas color.) ; 29 cm.

Inclui anexo.
Orientador: Elizabeth Pacheco Batista Fontes.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.90-100.

1. Soja - Melhoramento genético. 2. *Glycine max*. 3. Silenciamento gênico induzido por vírus. 4. Estresse no retículo endoplasmático. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 631.52

PRISCILA ALVES DA SILVA

VETOR DE SILENCIAMENTO GÊNICO E ANÁLISE COMPREENSIVA DA RESPOSTA AO ESTRESSE NO RETÍCULO ENDOPLASMÁTICO EM SOJA

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

APROVADA: 8 de maio de 2015

Carlos Priminho Pirovani

Anésia Aparecida dos Santos

Francisco Murilo Zerbini Júnior
(Co-Orientador)

Maximiller Dal-Bianco Lamas Costa

Elizabeth Pacheco Batista Fontes
(Orientadora)

Aos meus pais, Manoel e Lídia,
aos meus irmãos, Adriano, Lucas e Paulina,
aos meus sobrinhos, Mateus e Henrique
DEDICO.

AGRADECIMENTOS

A DEUS, minha fortaleza! Por me guiar, proteger, iluminar, abençoar meu caminho. A Nossa Senhora por sua intercessão em minha vida.

Aos meus pais, que não medem esforços para me permitir realizar meus sonhos. Muito obrigada por serem tão generosos, carinhosos, por me apoiarem, me darem forças e me incentivar e amparar nos momentos mais difíceis.

Aos meus irmãos, meus grandes amigos, não há palavras que descreva o tamanho do meu amor e gratidão por vcs e por nossos pais. AMO VCS!!! Obrigada, por poder contar sempre com vocês e, por terem me apoiado e incentivado tanto para concluir meu doutorado.

A minha cunhada Renata e ao meu cunhado Neto, pelo carinho e torcida.

Aos meus sobrinhos, Mateus e Henrique, que são verdadeiros anjinhos que chegaram a nossa família e nos enchem de alegrias.

A Lourdes, "minha segunda mãe", pelo imenso carinho e torcida.

A minha orientadora, Elizabeth Pacheco Batista Fontes, pela oportunidade, ensinamentos, exemplo de dedicação e empenho a ciência.

A Universidade Federal de Viçosa – UFV, pela excelente estrutura em ensino e pesquisa e ao Programa de Pós-Graduação em Genética e Melhoramento.

A CAPES e CNPq pelo apoio financeiro.

Aos membros da banca: Prof. Carlos Priminho Pirovani, Francisco Murilo Zerbini, Anésia Aparecida dos Santos e Prof. Maximiller Dal-Bianco Lamas Costa.

Ao pesquisador Francisco José Lima Aragão, pela disponilidade e atenção em me receber em seu laboratório e realização do ensaio de *Northen blot.*

Aos meus queridos amigos e mais que especiais do Laboratório de Biologia Molecular de Plantas - LBMP, se torna até difícil descrever e agradecer pelo excelente convívio, ensinamentos, amizade, carinho, apoio, ajuda e torcida durante todo meu doutorado. "LBMP: Prof. Elizabeth, João Paulo, Janaína, Giselle, Anésia, Iara, Pedro, Otávio, Cleysinho, Bianca, Maiana, Hanna, Claudinha, Lucas, Nivea, Welison, Michihito, Rafinha, Bruno, Vírgilio, Marco Aurélio, Laura, Marcia, Dani, Cris, Humberto, Fábia, Marlene, Gláucia, Adriano, Wellington, a todos vocês muito obrigada!!!"

A minha estagiária, Hanna, dedicada, muito obrigada por sua ajuda.

Dani e Iara, que iniciaram o trabalho com VIGS e sempre atenciosas comigo. Muito obrigada!

Ao Cleysinho, pela dedicação ao trabalho de bioinformática. Ao Otávio, pela contribuição no trabalho de bioinformática e estatística.

As amigas de república: Dalila, Camila, Marcele e Camila e as amigas: Dani e Rose, pela amizade, apoio e carinho.

Aos meus "amigos(as) distantes" mas que sempre torcem por mím.

A Comissão Permanente de Propriedade Intelectual da UFV, em especial Ivan, pela atenção e disponibilidade durante processo de patenteamento.

Ao prof. Maximiller pela ajuda e atenção na etapa final de conclusão da tese.

À Marlene, Glaucia, Fábia e Adriano pela presteza.

Ao Laboratório de Fisiologia Molecular de Plantas da UFV agradeço por permitir a realização das leituras de fluorescência, em especial a Ana, pela atenção.

Ao padre Sebastião, pelas missas abençoadas na capelinha da UFV.

A todos os funcionários do BIOAGRO/UFV, em especial Sr. Paulo sempre atencioso na portaria do BIOAGRO.

E a todos que de alguma forma direta ou indiretamente me ajudaram e me apoariam.

MUITO OBRIGADA!!! Jesus os abençoem!!!

SUMÁRIO

RESUMO

SILVA, Priscila Alves da, D.Sc., Universidade Federal de Viçosa, maio de 2015. **Vetor de silenciamento gênico e análise compreensiva da resposta ao estresse no retículo endoplasmático em soja.** Orientadora: Elizabeth Pacheco Batista Fontes. Coorientadores: Francisco Murilo Zerbini Júnior, Juliana Rocha Lopes Soares Ramos e Michihito Deguchi.

O silenciamento de RNA pode ser induzido em plantas como resultado da infecção por vírus, um processo denominado "silenciamento gênico induzido por vírus" (*virus induced gene silencing)* – VIGS. Vetores VIGS permitem de uma forma rápida e eficiente a análise funcional de genes de soja por genética reversa. A construção de um vetor de silenciamento viral para inativação de genes de soja foi baseado na modificação do genoma do DNA-A de SoCSV (*Soybean chlorotic spot virus*) de forma a deletar o gene da proteína do capsídeo, e incorporar sítios de enzimas de clonagem chaves que permitam a inserção de sequências para silenciamento específico de genes alvos. O vetor de silenciamento derivado do genoma de SoCSV poderá ser usado para permitir o entendimento entre as possíveis comunicações cruzadas entre a via UPR e a via de morte celular mediada por NRPs (DCD/NRP), já que a falta de conhecimento com relação aos transdutores de sinais da referida via UPR em soja tem limitado seu estudo. A via UPR (*Unfolded Protein Response*) é desencadeada pelo acúmulo de proteínas mal dobradas e, por sua vez, é induzida para garantir o dobramento adequado das proteínas atenuando o estresse no RE. Nos mamíferos, a via UPR muito bem caracterizada é regulada pelo chaperone molecular, BiP, e ativada por três receptores: PERK, IRE1 e ATF6. A caracterização da via UPR em Arabidopsis foi baseada em mamífero e é ativada por duas classes de receptores transmembranas do RE: bZIP28 e bZIP17 (homólogos de ATF6) e IRE homólogos, IRE1a e IRE1b. Apesar da importância do RE como uma organela chave envolvida em respostas adaptativas a estresse, a resposta ao estresse RE não tem sido caracterizada no genoma da soja. Portanto, por análise in silico, em comparação com Arabidopsis, e estudos funcionais foi gerado um painel com um completo cenário de resposta ao estresse no RE em soja. No genoma da soja foram também identificados fatores de transcrição induzidos por estresse no RE, um associado à membrana, ortólogo de AtNAC062 e um fator de transcrição ortólogo de AtNAC103. Além de genes envolvidos na via UPR, foi identificado um fator de transcrição associado à membrana do RE, ortólogo de AtNAC89, que contém um domínio NAC e up-regula

genes associados a morte celular. Em soja foi demonstrado que o sinal de estresse no RE e estresse osmótico integra com outras repostas adaptativas, como a via de sinalização de morte celular programada mediada por DCD/NPRs, específica de planta e iniciada por GmERD15 que ativa o promotor alvo NRP. A expressão aumentada de NRP leva à indução de GmNAC81 e GmNAC30 que cooperam para ativação da expressão do gene VPE (*vacuolar processing enzyme*), que, por sua vez, executa o processo de morte celular programada. Ao contrário, por análise *in silico* mostramos que esta via de morte celular mediada por DCD/NRP também é conservada em Arabidopsis, com exceção para GmERD15, que aparententemente não possui ortólogo em Arabidopsis. Em nossa pesquisa fornecemos evidências que as respostas de estresse no RE funcionam de forma semelhante. Primeiramente uma alta conservação de estruturas primárias de soja com preditos ortólogos de Arabidopsis possuem domínios de localização e funcional comuns que podem ser associados com atividade bioquímica correta e localização subcelular. Em segundo, os dois ramos da UPR em soja foram analisados funcionalmente. Ortólogos bZIP17/28 (GmbZIP37 e GmbZIP38) e ortólogo ZIP60 (GmbZIP68) de soja possuem organização estrutural similar e assim como em Arabidopsis são induzidos por estresse no RE e o putativo substrato de GmIREs, transcrito GmbZIP68, possui sítio canônico para atividade endonuclease de IRE1 e sofre splicing sob condições de estresse no RE. A expressão de bZIP38, bZIP37 e bZIP68 suporta que a via UPR em plantas é funcionalmente conservada em soja. O amplo painel compreensivo de resposta a estresse do RE em soja gerado permite predições funcionais da sinalização de componentes de estresse no RE. Portanto, o vetor VIGS desenvolvido possibilitará um avanço na caracterização funcional dos componentes da via e possibilidade de conecções com respostas a múltiplos estresses em plantas.

ABSTRACT

SILVA, Priscila Alves da, D.Sc., Universidade Federal de Viçosa, May, 2015. **Gene silencing vector and comprehensive analysis of the endoplasmic reticulum stress response in soybean.** Adviser: Elizabeth Pacheco Batista Fontes. Co-advisers: Francisco Murilo Zerbini Júnior, Juliana Rocha Lopes Soares Ramos and Michihito Deguchi.

The RNA silencing can be induced in plants as a result of virus infection, this process is called *virus induced gene silencing* – VIGS. VIGS vectors allow a fast and efficient functional analysis of soybean genes by reverse genetics. The construction of a VIGS vector for the soybean genes inactivation was based in the modification of the DNA-A of SoCSV (*Soybean chlorotic spot virus*) genome aiming to delete the protein gene of the capsid, and to incorporate enzyme sites of cloning keys that allow the insertion of sequences to the silencing of specific target genes. The silencing vector derived from the SoCSV genome can be used to enable a better understanding between the possible cross-communication between the UPR pathway and the pathway of cell death mediated by NRPs (DCD/NRP), since the lack of knowledge about transducers signals of the UPR via has limited its study. The UPR (Unfolded Protein Response) is triggers by the accumulation of unfolded proteins, and it is induced to guarantee the proper protein folding attenuating the stress on ER **(**endoplasmic reticulum**).** In mammals, a very well characterized UPR pathway is regulated by the molecular chaperone BiP, and it is activated by three receptors: PERK, IRE1 e ATF6. The UPR pathway characterization in Arabidopsis was based on mammals and it is activated by two classes of transmembrane receptors of ER: bZIP28 e bZIP17 (homologs of ATF6) and IRE homologs, IRE1a and IRE1b. Despite the importance of the RE as a key organelle involved in adaptative responses to stress, the ER stress response has not been characterized in the soybean genome. Therefore, by *in silico* analysis, in comparison with Arabidopsis and functional studies, a panel was created with a complete scenario of ER stress response in soybean. In the soybean genome were also identified transcription factors induced by stress in the ER, one of them associated to the membrane, orthologs of AtNAC062, and a transcription factor ortholog of AtNAC103. Besides the genes involved on the UPR pathway, a transcription factor associated with the ER membrane, orthologs of AtNAC89, that contains a NAC domain and up regulates genes associated to the cell death was identified. In soybean, it was demonstrated that the ER stress and osmotic

stress-signal integrates with other adaptive responses, as the plant-specific cell death signaling pathways mediated by DCD/NPRs and initiated by GmERD15, that activates the target promoter NPR. The increased expression of NPR leads to the induction of GmNAC81 and GmNAC30 that cooperate to the activation of the expression of the VPE gene (vacuolar processing enzyme), that executes the process of programmed cell death. In a reverse approach, by *in silico* analysis we also examined the Arabidopsis genome for components of a previously characterized ER stress-induced cell death signaling response in soybean. With the exception of GmERD15, which apparently does not possess an Arabidopsis ortholog, the Arabidopsis genome harbors conserved GmNRP, GmNAC81, GmNAC30 and GmVPE sequences that share significant structural and sequence similarities with their soybean counterparts. In our research we showed evidences that the stress response in the ER operates in a similar fashion in both soybean and Arabidopsis. Firstly, a high conservation of primary structures of soybean with predicted Arabidopsis orthologs have functional common domains that can be associated to the biochemistral activity and subcellular localization. Secondly, both branches of UPR in soybean were analysed functionally. The bZIP17/bZI28 orthologs (GmbZIP37 and GmbZIP38) and ZIP60 ortholog (GmbZIP68) from soybean have a similar structural organization and as in Arabidopsis they are induced by stress in ER and the putative substrate of GmIREs, transcript GmbZIP68, have canonical site to endonuclease activity of IRE1 and undergoes alternatively spliced under stress conditions in ER. The expression of bZIP38, bZIP37 and bZIP68 supports that the UPR pathway in plants is functionally conserved in soybean. Our *in silico* analyses, along with functional data, have generated a comprehensive overview of the ER stress response in soybean as a framework for functional prediction of ER stress signaling components. Therefore, the development of VIGS vector enables an advance in the functional characterization of the pathway components and the possible connections with multiple stress responses in plants.

INTRODUÇÃO GERAL

A soja é uma cultura agronomicamente de grande relevância no cenário agrícola nacional e internacional e perdas de produtividade desencadeadas por estresses bióticos ou abióticos, tais como alta salinidade, estresse hídrico, ataque por patógenos, podem gerar um prejuízo econômico muito significativo. Diante dessas condições adversas que podem alterar aspectos morfológicos, fisiológicos, bioquímicos e/ou moleculares, limitando o crescimento, desenvolvimento e a produtividade das plantas (Buchanan, 2000), pesquisas que visam ao entendimento de mecanismos de resposta das plantas aos estresses são de extrema importância para desenvolver cultivares mais adaptadas às condições de estresses ambientais.

Condições de estresses podem alterar a homeostase do retículo endoplasmático (RE), uma organela eucariótica que desempenha um papel central na maturação e transporte de proteínas secretórias, causando acúmulo de proteínas mal dobradas na organela que resulta na ativação de uma resposta adaptativa conservada em eucariotos conhecida como UPR (Unfolded Protein Response) (Ron e Walter, 2007). A ativação da UPR leva à redução da síntese geral de proteínas, aumento da capacidade de processamento do RE e ativação de mecanismos de degradação de proteínas mal dobradas, um conjunto de respostas coordenadas que visam restaurar a homeostases do RE em condições de estresses (Schroder e Kaufman, 2005).

O RE possui sensores de estresse, que são receptores integrais da membrana da organela que monitoram a função reticular e sinalização da UPR. Em células de mamíferos, a via UPR é muito bem caracterizada, sendo ativada pelos transdutores de sinais: PERK, IRE1 e ATF6. A ativação da via UPR é regulada pelo chaperone molecular, BIP. Em condições normais, BiP se associa ao domínio luminal desses três receptores e os mantém na forma inativa. Sob condições de estresse, BIP se dissocia dos mesmos, resultando na oligomerização e ativação de IRE1 e PERK, além da translocação de ATF6 para o complexo de Golgi. No complexo de Golgi, ATF6 sofre proteólise por S1P e S2P resultando em sua liberação e direcionamento para núcleo onde se liga aos promotores para ativar a expressão de genes alvo da UPR (Ye et al., 2000 Schindler et al., 2009). Já a ativação de PERK, mediada pela dissociação de BiP, promove a sua dimerização e consequente ativação do domínio cinase voltado para o citosol. A ativação da cinase PERK provoca a fosforilação do

fator de iniciação da tradução eIF2α, causando inibição da tradução global das proteínas (Harding et al., 2000) e indução de ATF4, outro fator de transcrição, que regula a expressão de genes envolvidos na ERQC (controle de qualidade do RE), metabolismo de aminoácidos, resistência ao estresse oxidativo e morte celular (Harding 2000 Jiang 2004). Por meio de um mecanismo de ativação similar, IRE1 quando ativado promove o splicing do mRNA específico que codifica o fator de transcrição bZIP, XBP-1, resultando na tradução de um fator transcricional funcional que regula positivamente a expressão de genes relacionados ao estresse no RE (Yoshida et al., 2001).

Os progressos recentes na elucidação da via UPR em plantas demonstraram que o sinal de estresse no RE é transduzido por duas classes de receptores transmembranas do RE: bZIP28 e bZIP17 (homólogos de ATF6) e IRE homólogos, Ire1a e Ire1b em Arabidopsis (Iwata e Koizumi, 2012 Howell, 2013 Reis e Fontes, 2013). Em plantas de Arabidopsis, IRE promove o splicing do mRNA que codifica o fator de transcrição bZIP60, deletando a região que codifica o domínio transmembrana, o que libera o fator tanscrional para o núcleo a fim de ativar os genes alvos da via UPR. O mecanismo de ativação de AtbZIP28 assemelha-se àquele descrito para ATF6 em mamíferos. Dissociação de BiP de AtbZIP28 induzida por estresses no RE possibilita a translocação do receptor da membrana para o Golgi onde sofre hidrólises proteolíticas específicas (Srivastava et al., 2013).

As análises de variação global de expressão gênica induzida por estresses no RE forneceram evidências que a via UPR funciona similarmente em soja (Isrigler et al., 2007). Os indutores de estresse no RE, tal como tunicamicina e AZC, induzem coordenadamente um conjunto de genes que funcionam no (i) dobramento de proteínas, (ii) no sistema de degradação de proteínas associado ao retículo endoplasmático (ERAD) e (iii) em regulação de tradução. Estas análises genômicas indicaram que soja, como Arabidopsis e mamíferos, evoluíram um mecanismo de ativação da UPR resultando em pelo menos duas respostas fisiológicas conservadas para recuperação do estresse no RE: (i) indução transcricional de genes que codificam chaperones moleculares e foldases e (ii) ativação do sistema ERAD para o descarte rápido de proteínas mal dobradas como parte do sistema de qualidade da organela. Entretanto, os receptores sensores e os mecanismos de propagação do sinal de estresse na via UPR não foram ainda identificados em soja.

Além de ativar a via UPR evolutivamente conservada, foi demonstrado em soja que o sinal de estresse no RE também comunica com outras repostas adaptativas, como a via de sinalização de morte celular mediada por NPRs e induzida por estresse no RE, estresse osmótico e seca (Irsigler et al., 2007 Costa et al., 2008 Valente et al., 2009). Estes estresses abióticos induzem a expressão do fator transcricional GmERD15 que ativa o promotor alvo NRP, induzindo a expressão do gene NRP. A expressão aumentada de NRP leva à indução de GmNAC81 e GmNAC30 que cooperam para ativação da expressão do gene VPE (vacuolar processing enzyme), que, por sua vez, executa o processo de morte celular programada (PCD Alves et al., 2011 Mendes et al., 2013 Reis e Fontes, 2012 e 2013). Os componentes da via de sinalização de morte celular mediada por NRP são também induzidos por outros estresses abióticos e bióticos (Faria et al., 2011 Pinheiro et al., 2009). É provável que esta via de sinalização que transduz um sinal de PCD represente uma resposta comum de células vegetais a múltiplos sinais de estresses.

Sendo um ramo da UPR que conecta com múltiplas respostas a estresses, a via de sinalização mediada por NRP deve permitir uma adaptação versátil de plantas a diferentes condições de estresses. Esta hipótese tem sido levantada após a observação de que a modulação da referida via pela superexpressão do chaperone molecular BiP confere às linhagens transgênicas uma melhor adaptação às condições de seca (Valente et al., 2009: Carvalho et al., 2014). Apesar do potencial dessa via de sinalização de morte celular como alvo para engenharia genética de tolerância a múltiplos estresses, muitos aspectos moleculares da referida via permanecem para serem elucidados. Por exemplo, ainda não foi identificado o receptor do RE que conecta molecularmente o sinal de estresse na organela à ativação da via de morte celular. Também não se conhece como o sinal de morte celular é propagado após a ativação de NRP até a indução dos genes GmNAC30 e GmNAC81. Uma explicação para o progresso limitado na elucidação das bases moleculares dessa resposta adaptativa refere-se à ausência de um sistema eficiente de silenciamento gênico em soja que permita os estudos de genética reversa para designar funções gênicas e ordenar componentes em vias de transdução de sinais. Além disso, a falta de conhecimento com relação aos transdutores de sinais da via UPR em soja tem limitado o nosso entendimento entre as possíveis comunicações cruzadas entre a via UPR e a via de morte celular mediada por NRPs. Sendo assim, nossa investigação

está estruturada com os objetivos principais: no capítulo I de desenvolver ferramentas moleculares para estudos funcionais de genes em soja através da construção de um vetor de silenciamento gênico induzido por vírus (VIGS) e no capítulo II caracterizar de uma forma compreensiva *in silico* e funcionalmente a via UPR em soja.

CAPITULO I

VETOR VIRAL E USO DESTE PARA SILENCIAMENTO GÊNICO EM PLANTAS LEGUMINOSAS

Patente: BR 10 2015 007046 2

Elizabeth Pacheco Batista Fontes, Priscila Alves da Silva, Daniela Coco, Iara Pinheiro Calil, Hanna Durso Neves Caetano

CÓPIA

INPI INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL PROTOCOLO GERAL
jfenatocg
30/03/2015 014150000510 12:40 DEMG
BR 10 2015 007046 2

< Uso exclusivo do INPI >

Espaço reservado para o protocolo | Espaço reservado para a etiqueta | Espaço reservado para o código QR

INPI INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
Sistema de Gestão da Qualidade
Diretoria de Patentes

DIRPA	Tipo de Documento: Formulário	DIRPA	Página: 1/3
Título do Documento: Depósito de Pedido de Patente		Código: FQ001	Versão: 01
		Procedimento: DIRPA-PQ006	

Ao Instituto Nacional da Propriedade Industrial:
O requerente solicita a concessão de um privilégio na natureza e nas condições abaixo indicadas:

1. Depositante (71):

1.1 Nome: UNIVERSIDADE FEDERAL DE VIÇOSA
1.2 Qualificação: FUNDAÇÃO
1.3 CNPJ/CPF: 25.944.455/000196
1.4 Endereço Completo: Av. P.H. Rolfs, s/n,Campus da UFV
1.5 CEP: 36570-000
1.6 Telefone: 31 3899 1421 1.7 Fax: 31 3899 2148
1.8 E-mail: propriedadeintelectual@ufv.br

☒ continua em folha anexa

2. Natureza: ☒ Invenção ☐ Modelo de Utilidade ☐ Certificado de Adição

3. Título da Invenção ou Modelo de Utilidade (54):

VETOR VIRAL E USO DESTE PARA SILENCIAMENTO GÊNICO EM PLANTAS LEGUMINOSAS

☐ continua em folha anexa

4. Pedido de Divisão: do pedido Nº **Data de Depósito:**

5. Prioridade: ☐ Interna (66) ☐ Unionista (30)

O depositante reivindica a(s) seguinte(s):

País ou Organização do depósito	Número do depósito (se disponível)	Data de depósito

☐ continua em folha anexa

INPI INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL PROTOCOLO GERAL
jfenatocg
30/03/2015 014150000510 12:40 DEMG
52400.010663/2015-18

6. Inventor (72):

☐ Assinale aqui se o(s) mesmo(s) requer(em) a não divulgação de seus nome(s), neste caso não preencher os campos abaixo.

6.1 Nome: Elizabeth Pacheco Batista Fontes

6.2 Qualificação: Professor

6.3 CPF: 332.896.546-72

6.4 Endereço Completo: Condomínio Recanto da Serra, 36, Viçosa, MG

6.5 CEP: 36570-000

6.6 Telefone: 31 38992949 6.7 FAX:

6.8 E-mail: bbfontes@ufv.br

☒ continua em folha anexa

7. Declaração de divulgação anterior não prejudicial. ☐
Artigo 12 da LPI – período de graça.
Informe no item 11.13 os documentos anexados, se houver.

8. Declaração na forma do item 3.2 da Instrução Normativa nº 17/2013:

☐ Declaro que os dados fornecidos no presente formulário são idênticos ao da certidão de depósito ou documento equivalente do pedido cuja prioridade está sendo reivindicada.

9. Procurador (74):

9.1 Nome: Afonso Sérgio Corrêa de Faria

9.2 CNPJ/CPF: 137.424.346-91 9.3 API/OAB: 21.972

9.4 Endereço Completo: Procuradoria Jurídica, Campus da UFV, Viçosa, MG.

9.5 CEP: 36570-000

9.6 Telefone: 31 3899 2110 9.7 FAX:

9.8 E-mail: prj@ufv.br

☐ continua em folha anexa

10. Listagem de sequências biológicas.
Informe nos itens 11.9 ao 11.12 os documentos anexados, se houver. ☐

11. Documentos Anexados:

(Assinale e indique também o número de folhas):
(Deverá ser indicado o número total de somente uma das vias de cada documento).

		Documentos Anexados	folhas
☒	11.1	Guia de Recolhimento da União (GRU).	1
☒	11.2	Procuração.	1
☐	11.3	Documentos de Prioridade.	
☐	11.4	Documento de contrato de trabalho.	
☒	11.5	Relatório descritivo.	30
☒	11.6	Reivindicações.	2
☒	11.7	Desenho(s) (se houver). Sugestão de figura a ser publicada com o resumo: nº, _____ por melhor representar a invenção (sujeito à avaliação do INPI).	6
☒	11.8	Resumo.	1
☒	11.9	Listagem de sequências em arquivo eletrônico: 2 nº de CDs ou DVDs (original e cópia).	
☒	11.10	Código de controle alfanumérico no formato de código de barras referente às listagem de sequências.	1
☐	11.11	Listagem de sequências em formato impresso.	
☒	11.12	Listagem de sequências - Declaração de acordo com a Resolução INPI nº 70/2013.	1
☒	11.13	Outros (especificar) DOU, DOE, Anexo	4

12. Total de folhas anexadas: 47 **fls.**

13. Declaro, sob as penas da Lei que todas as informações acima prestadas são completas e verdadeiras.

Viçosa, 27/04/2015
Local e Data

Assinatura e Carimbo

Afonso Sérgio Côrrea de Faria
PROCURADOR FEDERAL/UFV
OAB/MG 21.972

VETOR VIRAL E USO DESTE PARA SILENCIAMENTO GÊNICO EM PLANTAS LEGUMINOSAS

CAMPO DA INVENÇÃO

[1] A presente invenção refere-se a um vetor de silenciamento viral baseado na modificação do genoma do vírus *Soybean chlorotic spot virus* (SoCSV) para induzir o silenciamento sistêmico de genes endógenos de plantas pertencentes à família *Fabaceae*, especialmente a soja, o que permite a análise funcional de genes de uma forma rápida e eficiente. Tal tecnologia possibilita um avanço no conhecimento básico de funções gênicas desconhecidas e na consequente seleção de genes de interesse para o melhoramento genético desta cultura de elevada importância econômica no Brasil.

ESTADO DA TÉCNICA

[2] Silenciamento gênico em plantas refere-se à supressão de um gene endógeno pela introdução de uma sequência de DNA exógena, porém relacionada, ou a supressão de transgenes inseridos em posições ectópicas no genoma. Eventos tanto citoplasmáticos quanto nucleares são implicados no silenciamento gênico que pode ser classificado em (i) silenciamento gênico pós-transcricional (post-transcriptional gene silencing, PTGS) e (ii) silenciamento gênico transcricional (TGS). PTGS é um mecanismo pelo qual a planta degrada de forma específica um determinado mRNA, bloqueando a expressão do gene sem afetar sua taxa de transcrição (Baulcombe D. RNA silencing in plants. 2004. Nature. 431, 356-363). Existem seis vias de silenciamento de RNA (Chapman E. J., Carrington J. C. Specialization and evolution of endogenous small RNA pathways. 2007. Vol 8, 884-896), sendo que a via do siRNA (*short interfering* RNA) tem um importante papel na defesa contra vírus e transposons. Na via de silenciamento do siRNA, a resposta antiviral da planta é iniciada quando a proteína DCL (RNase III DICER-Like1), com o auxilio de HYL1 (HYPONASTIC LEAVES1), cliva o RNA viral fita dupla (dsRNA) em pequenos RNAs designados siRNAs, de 21 a 24 nucleotídeos, no citoplasma da planta. Uma única fita de RNA, incorporada ao RISC (RNA-*induced silencing complex*), do qual AGO1 (argonauta) faz parte, guia o reconhecimento e degradação de RNAs complementares ao RNA que gerou os siRNAs (Ding S.W., Voinnet O. Antiviral immunity directed by small RNAs. 2007. 130, 413-426). Uma

vez que a própria replicação viral ativa o silenciamento de RNA, a inserção no genoma viral de um fragmento de um gene não silenciado (endógeno ou transgene) irá levar ao silenciamento do gene correspondente na planta. Esta forma de induzir o silenciamento por meio da replicação viral foi denominada Silenciamento Gênico Induzido por Vírus (*virus-induced gene silencing*, VIGS Kumagai M.H., Donson J., Della-Cioppa G., Harvey D., Hanley K., Grill L.K. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. 1995. Proc. Natl. Acad. Sci. 92, 1679-1683 Ruiz M.T., Voinnet O., Baulcombe D.C. Initiation and maintenance of virus-induced gene silencing. 1998. Plant Cell. 10, 937).

[3] O uso de VIGS para o estudo da expressão e função de genes constitui atualmente uma das aplicações práticas mais disseminadas do silenciamento de RNA (Lu R., Martin-Hernandez A.M., Peart J.R., Malcuit I., Baulcombe D.C. Virus-induced gene silencing in plants. 2003. Methods. 30, 296–303). Para indução de VIGS, um fragmento do gene de interesse é inserido em um vetor viral, em substituição a genes dispensáveis para a infecção sistêmica ou como um gene extra. Por meio de interação RNA:RNA, o vetor para VIGS provoca o silenciamento do gene alvo específico. Sendo assim, VIGS tem sido utilizado como uma importante ferramenta na análise funcional de genes em plantas (Burch-Smith T.M., Anderson J.C., Martin G.B., Dinesh-Kumar S.P. Applications and advantages of virus-induced gene silencing for gene function studies in plants. 2004. Plant Journal. 39, 734-746). Esta é uma técnica que oferece grandes vantagens por ser mais rápida e permitir a análise de genes em larga escala. Os vetores para VIGS têm sido frequentemente empregados em estudos de genética reversa, sendo promissores para plantas difíceis de serem transformadas, caso estas sejam suscetíveis à infecção viral (Robertson D. VIGS vectors for gene silencing: many targets, many tools. 2004. Annu. Ver. Plant Biol. 55, 495–519). Além disso, VIGS permite analisar genes que possuem um fenótipo letal, uma vez que essa técnica é aplicada em plantas adultas, não havendo morte do embrião como ocorre em sistemas de mutagênese insercional (Baulcombe D.C. Fast forward genetics based on virus-induced gene silencing. Curr. Opin. 1999. Plant. Biol. *2*, 109-113).

[4] Atualmente, a análise da função dos genes utilizando VIGS é realizada em diversas culturas e inúmeros genomas virais já foram manipulados para funcionarem como vetores para VIGS. Em tabaco, tomate, pimenta e Arabidopsis thaliana, o vetor de silenciamento baseado na modificação do genoma do vírus de

RNA Tobacco ratle virus (TRV) é o mais utilizado (Ratcliff, Frank. Tobacco rattle virus as a vector for analysis of gene function by silencing. 2001 Vol 25, 237-245), porém este vetor é limitado para o Brasil, pois o TRV é uma praga quarentenária ausente em nosso país. Outros vetores para VIGS utilizados incluem vírus de RNA, como *Potato virus* X (PVX), Bean pod mottle virus (BPMV US 2007/0214518), *Pea early browning virus* (PEBV C G.D., Krath B.N., MacFarlane S.A., Nicolaisen M., Johansen I.E., Lund O.S. Virus-induced gene silencing as a tool for functional genomics in a legume species. 2004. Plant J. 40, 622–31) e Soybean yellow mottle mosaic virus (SYMMV KR 2012/053710). Entre eles, Bean pod mottle virus (BPMV) e Soybean yellow mottle mosaic virus (SYMMV) têm sido utilizados para desenvolvimento de vetores para VIGS para análise funcional de genes em soja. Entretanto, estes vetores para VIGS são derivados de genomas de vírus de RNA que são suscetíveis ao RNAi, uma vez que eles possuem um genoma de RNA que replica no citoplasma, o que compromete o silenciamento estável de um gene alvo de soja por um período mais longo.

[5] Os vírus de DNA para desenvolvimento de vetores para VIGS que têm sido utilizados são os begomovírus Tomato golden mosaic virus (TGMV Kjemtrup S., Sampson K.S., Peele C.G., Nguyen L.V., Conkling M.A. Gene silencing from plant DNA carried by a Geminivirus. 1998. Plant J. 14, 91–100), Cabbage leaf curl virus (CaLCuV Turnage M.A., Muangsan N., Peele C.G., Robetson D. Geminivirus-based vectors for gene silencing in Arabidopsis. 2002. Plant Journal 30, 107-114) e African cassava mosaic virus (ACMV Fofana I., Sangare A., Collier R., Taylor C., Fauquet C. A geminivirus-induced gene silencing system for gene function validation in cassava. 2004. Plant Mol Biol. 56, 613–24). Os begomovírus replicam no núcleo de células infectadas, como episssomos, e, se carregando sequências homólogas de DNA, podem modular a expressão de genes ativos nos cromossomas de células vegetais, referidos como genes endógenos. Entretanto, nenhum dos vetores para VIGS descritos que foram construídos a partir de modificação do genoma de vírus de DNA (begomovírus), pode ser utilizado para estudos funcionais de genes de soja, uma vez que são limitados na sua capacidade de infectar leguminosas.

[6] Os vírus de plantas do gênero begomovírus pertencem à família *Geminiviridae*. A família *Geminiviridae* é constituída por sete gêneros *Mastrevirus*, *Curtovirus*, *Begomovirus*, *Topocuvirus*, *Becurtovírus*, *Eragrovírus* e *Turncurtovírus* (Bernardo P., Golden M., Akram M., Naimuddin M., Nadarajan N., Fernandez E.,

Granier M., Rebelo A.G., Peterschmitt M., Martin D.P., Roumagnac P. Identification and characterisation of a highly divergent geminivirus: evolutionary and taxonomic implications. 2013. Virus Research. 177, 35-45 Varsani A., Navas-Catillo J., Mariones E., Hernández-Zepeda C., Idris A., Brown J.K., Zerbini F.M., Martin D.P. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. 2014. Arch. Virol. 159, 2193–2203). O gênero begomovírus apresenta o maior número de espécies da família *Geminiviridae.*

[7] Os vírus de plantas classificados na família *Geminiviridae* são caracterizados pelo genoma de DNA circular de fita simples (ssDNA) com aproximadamente 2.500 a 3.000 nucleotídeos, encapsidado por uma única proteína capsidial (CP). A CP se arranja na forma de 22 capsômeros formando dois icosaedros incompletos geminados, com cerca de 18 x 30 nm (Stanley J., Bisaro D.M., Briddon R.W., Brown J.K., Fauquet C.M., Harrison B.D., Rybicki E.P., Stenger, D.C. Family *Geminiviridae*. In: FAUQUET C.M., MAYO M.A., MANILOFF J., DESSELBERGER U., BALL, L.A. (Ed.). Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. 2005. San Diego: Elsevier Academic Press. p. 301-326).

[8] O gênero begomovírus possui espécies que apresentam um ou dois componentes genômicos, e são transmitidas pelos diferentes biótipos de *Bemisia tabaci* a plantas dicotiledôneas (Stanley J., Bisaro D.M., Briddon R.W., Brown J.K., Fauquet C.M., Harrison B.D., Rybicki E.P., Stenger, D.C. Family *Geminiviridae*. In: FAUQUET C.M., MAYO M.A., MANILOFF J., DESSELBERGER U., BALL, L.A. (Ed.). Virus Taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses. 2005. San Diego: Elsevier Academic Press. p. 301-326). Os begomovírus, pertencentes ao "Velho Mundo" (Europa, Ásia e África), apresentam em sua maioria um componente genômico (monossegmentados) e os encontrados no "Novo Mundo" possuem genoma dividido em dois componentes, denominados DNA-A e DNA-B, ambos com aproximadamente 2600 nucleotídeos. No DNA-A, encontram-se os genes codificadores das proteínas responsáveis pela replicação viral e encapsidação. O DNA-B possui genes codificadores das proteínas de movimento célula-a-célula e longa distância do vírus na planta (Rojas M.R., Hagen C., Lucas W.J., Gilbertson, R.L. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. 2005. Annual Review of Phytopathology. 43, 361-394). Ambos os componentes são necessários para a infecção sistêmica de plantas e com

exceção da região intergênica de aproximadamente 200 nucleotídeos, denominada região comum (RC), os dois componentes não apresentam identidade em suas seqüências de nucleotídeos.

[9] Os begomovírus estão entre os vírus mais bem caracterizados do ponto de vista molecular. Os atributos que permitiram o progresso na caracterização molecular e funcional do genoma viral incluem:

(a) genoma de DNA de tamanho reduzido

(b) replicação por meio de um intermediário de DNA fita dupla (dsDNA), facilmente manipulável pelos métodos padrão de clonagem

(c) desenvolvimento de métodos de inoculação independentes do inseto vetor.

[10] Estas características têm feito com que os begomovírus sejam utilizados como uma ferramenta para estudo da replicação do DNA e da regulação da expressão gênica em plantas mono e dicotiledôneas (Hanley-Bowdoin L., Settlage S.B., Robertson D. Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. 2004. Molecular Plant Pathology. 5,149-156 Rojas M.R., Hagen C., Lucas W.J., Gilbertson, R.L. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. 2005. Annual Review of Phytopathology. 43, 361-394). Pelo fato de os begomovírus produzirem um grande número de cópias de replicons extracromossomais (episssomas), o potencial de utilização destes para a expressão ou silenciamento de genes endógenos das plantas também tem gerado interesse considerável (Robertson D. VIGS vectors for gene silencing: many targets, many tools. 2004. Annu. Ver. Plant Biol. 55, 495–519).

[11] Apesar da existência de vetores de silenciamento funcionais derivados de genomas virais, existem poucos vetores eficientes para a inoculação em leguminosas, como plantas de soja e feijoeiro (Senthil-Kumar M., Gowda H.V.R., Hema R., Mysore K.S., Udayakumar M. Virus-induced gene silencing and its application in characterizing genes involved in water-deficit-stress tolerance. 2008. Journal of Plant Physiology. 165. 1404-1421). Torna-se necessário o estudo de novas espécies de vírus para serem utilizados como vetores de silenciamento viral em soja.

[12] O *Soybean chlorotic spot virus* (SoCSV) foi caracterizado em 2012 como uma nova espécie de *Begomovirus* que infecta soja e feijoeiro (Coco D., Calil I.P., Brustolini O. J. B., Santos A.A., Inoue-Nagata A.K., Fontes E.P.B. Soybean chlorotic spot virus, a novel begomovirus infecting soybean in Brazil. 2013. Arch

Virol. 158, 457–462). Por não desenvolver sintomas severos à cultura da soja, SoCSV pode ser utilizado para construção de um vetor para VIGS baseado na modificação de seu genoma. É de interesse geral para biotecnologia e melhoramento de soja, estudar o efeito do silenciamento de genes a partir do desenvolvimento de um vetor não convencional para VIGS, baseado na modificação do genoma de um vírus de DNA, como o SoCSV que infecta soja e feijoeiro.

[13] Como a maioria dos begomovírus encontrados no Brasil, SoCSV é bissegmentado, ou seja, possui dois componentes genômicos, denominados componentes A e B, ambos com aproximadamente 2.600 nucleotídeos. Os componentes genômicos de SoCSV possuem uma configuração típica do DNA-A e DNA-B de begomovírus (Hanley-Bowdoin L., Bejarano E.R., Robertson D. Mansoor S. Geminiviruses: masters at redirecting and reprogramming plant processes. 2013. Nat. Reviews 11:777-788). O DNA-A contém cinco genes (*CP*, *Rep*, *TrAP*, *REn* e *AC4*), que codificam para as funções requeridas na replicação do DNA (*Rep, replication initiator protein* e *REn, replication enhancer protein*), controle da expressão gênica (*TrAP, transcriptional activator protein*), supressão das defesas do hospedeiro (*TrAP* e AC4) e encapsidação (*CP*, *coat protein*). O DNA-B contém os genes que codificam as proteínas de movimento NSP (*nuclear shuttle protein*) e MP (*movement protein*), na fita viral e complementar, respectivamente (Coco D., Calil I.P., Brustolini O. J. B., Santos A.A., Inoue-Nagata A.K., Fontes E.P.B. Soybean chlorotic spot virus, a novel begomovirus infecting soybean in Brazil. 2013. Arch Virol. 158, 457–462). Estas proteínas estão envolvidas com o movimento do vírus durante a infecção e afetam a patogenicidade viral (Hanley-Bowdoin L., Bejarano E.R., Robertson D. Mansoor S. Geminiviruses: masters at redirecting and reprogramming plant processes. 2013. Nat. Reviews 11:777-788). Ambos os componentes são necessários para a infecção sistêmica de plantas. Com exceção da região intergênica de aproximadamente 200 nucleotídeos, denominada região comum (RC), os dois componentes não apresentam identidade em suas seqüências de nucleotídeos. A RC contém a origem de replicação (*ori*) dos geminivírus, que inclui diversos elementos de sequência altamente conservados entre as diferentes espécies do gênero e que são reconhecidos pela proteína de replicação viral Rep (Arguello-Astorga G., Guevara-Gonzalez R.G., Herrera-Estrella L.R., Rivera Bustamante R.F. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. 1994. Virology. 203, 90-100 Fontes E.P.B., Eagle

P.A., Sipe P.S., Luckow V.A., Hanley-Bowdoin L. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. 1994. Journal of Biological Chemistry. 269, 8459-8465).

[14] O desenvolvimento de um vetor de silenciamento viral baseado na modificação do genoma do vírus SoCSV para induzir o silenciamento sistêmico de genes endógenos de plantas pertencentes à família *Fabaceae*, especialmente soja, permitirá de uma forma rápida e eficiente a análise funcional de genes. Isto se torna particularmente relevante, uma vez que o genoma da soja foi recentemente sequenciado. O conhecimento da sequência do genoma da soja forneceu informações robustas com relação aos tipos de genes presentes, mas a função precisa da maioria desses genes não foi caracterizada. A determinação da função gênica é crucial para melhoramento genético de cultivares de soja por meio de melhoramento clássico ou engenharia genética. A construção de um vetor viral de plantas, baseado na modificação do genoma do vírus de DNA que infecta cultivares de soja, é extremamente útil no estudo da expressão de genes de soja, e permitirá um avanço no conhecimento básico de funções gênicas desconhecidas e na consequente seleção de genes de interesse para o melhoramento genético desta cultura de elevada importância econômica para o Brasil. A sociedade, o público em geral e a mídia estão aceitando melhor a engenharia genética usando sequências de DNA ou genes da mesma planta ou de plantas relacionadas que podem ser usadas para cruzamentos genéticos, o que torna o entendimento da função de genes de soja ainda mais importante. Para realmente explorar o potencial do genoma da soja como fundamento para alterar vias metabólicas e aumentar propriedades nutricionais e agronômicas é essencial a elucidação da função de seus genes.

[15] Devido à dificuldade em transformar soja, torna-se irreal usar a tecnologia de RNA de interferência (RNAi) como transgene para determinar função de genes.

[16] A presente invenção usa um vírus de DNA resultando em tecnologia similar ao RNAi em um processo designado Silenciamento Gênico Induzido por Vírus (VIGS). VIGS não requer transformação de plantas e produz resultados mais rápidos do que a transformação *per si* (Robertson D. VIGS vectors for gene silencing: many targets, many tools. 2004. Annu. Ver. Plant Biol. 55, 495–519). Outros vírus têm sido adaptados para VIGS em soja, mas são baseados em vírus de RNA, que são suscetíveis ao RNAi, uma vez que eles possuem um genoma de RNA

que replica no citoplasma. Os vírus de DNA, como SoCSV, replicam no núcleo de células vegetais e seus genomas não são susceptíveis ao silenciamento por RNAi. Isto faz com que vetores de silenciamento baseados em vírus de DNA resultem em um processo mais estável de silenciamento de um gene alvo de soja por um período mais longo. Esta propriedade é especialmente importante para caracteres reprodutivos, uma vez que a maioria das plantas requer que o vetor para VIGS seja inoculado em estágios de plântulas.

[17] Por meio da inoculação da planta com tal vetor de silenciamento, contendo um fragmento idêntico ou com grande semelhança a sequências de um gene endógeno da planta, é possível silenciar sistemicamente a expressão de um gene endógeno na planta, bem como investigar a função de tal gene na planta.

[18] A presente invenção proporciona um método de silenciamento de genes endógenos de plantas utilizando um DNA episomal, e fornece uma construção de um vetor de DNA viral para a utilização de tal método.

[19] A presente invenção proporciona um método de rastreio rápido e reprodutível de genes de plantas com função desconhecida, para determinar a sua função no tecido da planta ou em plantas intactas. Tal método de rastreio inclui a inserção no vetor de silenciamento de um DNA heterólogo idêntico ou com alta similaridade de sequência com a sequência do gene alvo inoculação das plantas hospedeiras com a construção com o intuito de induzir o silenciamento e, após um período de crescimento, a comparação da planta hospedeira inoculada, com uma planta controle não infectada. Construções com base em genoma de geminivírus são particularmente úteis, uma vez que estes vírus infectam um grande número de variedade de plantas economicamente importantes. As características de comparação entre plantas de teste e controle incluem: crescimento, morfologia, fenótipo observável, e composição bioquímica. As diferenças entre as plantas de teste e controle indicam a possível função da sequência de DNA alvo do silenciamento. O período de crescimento necessário para se tornar aparente qualquer diferença entre as plantas tratadas e controle irá variar dependendo das plantas hospedeiras utilizadas e da função do DNA a ser suprimida. Esses períodos podem variar de alguns dias, poucas semanas, até seis meses ou mais. Como o presente método não exige cultura de tecidos ou seleção para que alterações na expressão gênica sejam obtidas, o método pode ser adaptado para automação em larga escala de rastreio de sequências de genes com função desconhecida em plantas. O presente método também pode ser

útil no rastreio rápido e reprodutível de porções de um gene de planta isolado com função conhecida, para identificar aquelas porções ou fragmentos de genes que são efetivos na prevenção ou supressão da expressão.

SUMÁRIO DA INVENÇÃO

[20] A presente invenção refere-se a um vetor para silenciamento compreendendo um genoma modificado de *Soybean Chlorotic spot virus* (SoCSV) capaz de conter um DNA heterólogo, em que o DNA heterólogo é idêntico ou tendo uma grande semelhança de sequências a um gene endógeno ou consiste de um fragmento de um gene endógeno de uma planta da família *Fabaceae*.

[21] Outro aspecto adicional da presente invenção é a construção de DNA compreendendo um genoma modificado do SoCSV, no qual foi inserido um sítio múltiplo de clonagem com sítios para enzimas de restrição que facilitam a clonagem de fragmentos de DNA exógenos sob o controle do promotor do gene que codifica a proteína capsidial (CP).

[22] Outro aspecto adicional da presente invenção é uma construção de DNA compreendendo, na direção 5’ para 3', uma origem de replicação de SoCSV DNA que codifica as proteínas necessárias para a replicação da construção de DNA e um segmento de DNA heterólogo, em que este DNA heterólogo é idêntico ou com grande semelhança a sequências de um gene endógeno de plantas da família Fabaceae.

[23] O uso desse vetor para o silenciamento de genes de plantas da família *Fabaceae* permite de uma forma rápida e eficiente a análise funcional de genes.

DESCRIÇÃO DETALHADA DA INVENÇÃO

[24] *Soybean Chlorotic spot virus* (SoCSV) é uma espécie do gênero begomovírus, infecta soja e feijoeiro com alta eficiência e replica como episssomos no núcleo de células infectadas. O genoma de SoCSV foi utilizado para a construção de um vetor para silenciamento gênico viral.

[25] No caso de vetores para VIGS baseados em begomovírus, o DNA episssômico (replicons extracromossomais) deve transportar um fragmento ou uma sequência de DNA homóloga à do gene endógeno a planta a ser silenciado. O episssoma de DNA deve ser capaz de se replicar em múltiplas cópias nos núcleos de

células das plantas, onde o silenciamento sistêmico é desejado. O episssoma deve ser capaz de mover-se célula-a-célula na planta ou induzir o movimento de um fator de supressão difusível, a fim de entrar e afetar as células distantes do ponto inicial de inoculação. O silenciamento do gene pode resultar em um fenótipo alterado, o que inclui alterações nas características que podem ser observadas visualmente (por exemplo, cor), nas medidas (altura ou características de crescimento) ou nas características bioquímicas (por exemplo, quantidades de produtos de genes alvo, incluindo RNA, proteínas ou peptídeos, ou produtos das vias metabólicas).

[26] Um gene endógeno de planta refere-se a um gene integrado no DNA cromossômico do genoma da planta. Genes endógenos incluem aqueles que ocorrem naturalmente no genoma da planta, bem como aqueles artificialmente introduzidos (por transformação mediada por *Agrobacterium* ou bombardeamento biobalístico, por exemplo).

[27] Os termos "silenciados" ou "silenciamento do gene" referem-se a uma redução dos níveis de expressão de um gene alvo. O silenciamento pode ocorrer ao nível transcricional ou pós-transcricional. O silenciamento pode ser completo, em que nenhum produto do gene é produzido, ou parcial, em que uma redução substancial na quantidade de produto do gene ocorre.

[28] O termo "silenciamento sistêmico" refere-se ao silenciamento de genes em plantas ou tecidos de plantas, onde o silenciamento do gene ocorre em células que são distantes do local da inoculação inicial do DNA episssomal.

[29] O termo “vetor de silenciamento” refere-se a uma construção de DNA capaz de se replicar dentro de uma célula hospedeira, autonomamente ou como uma parte integrada do genoma do hospedeiro, transportando uma sequência de DNA heteróloga que é semelhante ou idêntica às sequências de nucleotídeos de um gene endógeno da planta hospedeira, ou a um fragmento de gene da planta. A sequência de DNA heteróloga é de similaridade de sequência suficiente para induzir o silenciamento do gene endógeno após a introdução do episssoma. As sequências de DNA heterólogas são altamente semelhantes à sequência de nucleotídeos dos genes endógenos para levar ao silenciamento.

[30] O termo "DNA heterólogo" se refere ao DNA contido na construção de DNA episssomal que não é naturalmente encontrado em conjunção com a construção de DNA episssomal, ou seja, DNA que tenha sido introduzido por técnicas de engenharia genética. O DNA heterólogo pode ser idêntico em sequência, ou ter alta

similaridade de sequência com o gene alvo a ser silenciado, sendo o fragmento heterólogo de um tamanho suficiente para induzir o silenciamento do gene endógeno. A presente invenção compreende um DNA episomal com base em um genoma viral.

[31] A presente construção, baseada no genoma do SoCSV, quando transferida para uma célula de planta, age induzindo o silenciamento de um gene já presente na célula da planta. O gene a ser silenciado pode ser um gene de planta endógeno, ou uma sequência de um gene ou de DNA que tenha sido artificialmente introduzido na célula da planta.

[32] Técnicas de bombardeamento de partículas podem ser usadas para introduzir o vetor de silenciamento em uma célula, ou em um grupo de células de uma planta. A replicação da construção na célula da planta conduz ao silenciamento sistêmico.

[33] As plantas que podem ser utilizadas na prática pela presente invenção são as espécies da família *Fabaceae*, onde se deseja o silenciamento sistêmico, e onde ocorra o movimento da construção de DNA. O termo "susceptível à infecção" inclui plantas que são naturalmente infectáveis por vírus, que podem ser inoculadas com a construção de DNA da presente invenção por métodos de inoculação mecânica. "Susceptível à infecção" refere-se ainda às plantas nas quais a construção de DNA é capaz de se replicar dentro da célula da planta inoculada.

[34] Uma variedade de técnicas está disponível para a introdução de construções de DNA numa célula de planta. Estas incluem transferência mediada por *Agrobacterium*, injeção, eletroporação, bombardeamento de micropartículas, etc. Para a presente invenção, as plantas foram inoculadas utilizando o bombardeamento de micropartículas. Microprojéteis transportando a construção de DNA de acordo com a presente invenção, incluindo tanto os componentes A e B do DNA, foram inoculados nas plantas, conforme o Experimento de Demonstração.

Construção do vetor para silenciamento gênico induzido por vírus (VIGS).

[35] As sequências de nucleotídeos completas de cada componente genômico de SoCSV, DNA-A, identificado como SEQ ID NO:17, [correspondente ao clone identificado pelo acesso pUFV1587 do Laboratório de Biologia Molecular de Plantas da Universidade Federal de Viçosa, (que se refere a uma cópia do DNA-A de SoCSV clonado no plasmídeo pBluescript KS+, identificado pelo acesso NC-018457.1

disponível no NCBI Coco, D., Calil, I.P., Brustolini, O. J. B., Santos, A.A., Inoue-Nagata, A.K., Fontes, E.P.B.. Soybean chlorotic spot virus, a novel begomovirus infecting soybean in Brazil. 2013. Arch Virol. 158, 457–462)], e DNA-B, identificado como SEQ ID NO:18, [correspondente aos clones identificados pelo acesso pUFV1588, (que se refere a uma cópia do DNA-B de SoCSV clonado no plasmídeo pBluescript KS+, identificado pelo acesso NC-018456.1 disponível no NCBI Coco, D., Calil, I.P., Brustolini, O. J. B., Santos, A.A., Inoue-Nagata, A.K., Fontes, E.P.B.. Soybean chlorotic spot virus, a novel begomovirus infecting soybean in Brazil. 2013. Arch Virol. 158, 457–462) e pUFV2571, (que se refere a uma cópia do DNA-B de SoCSV clonado no plasmídeo pBluescript KS+ do Laboratório de Biologia Molecular de Plantas da Universidade Federal de Viçosa)], foram submetidas à análise de restrição, utilizando-se o programa *ApE- A Plasmid Editor* para análise e construção de um clone infeccioso. De acordo com o resultado da análise de restrição gerada, foram selecionadas enzimas que clivaram o DNA viral liberando apenas um fragmento que contém a região comum de replicação, e este produto de clivagem foi clonado no plasmídeo pBluescript KS+, resultando na construção da "meia-cópia".

[36] Ressalta-se que qualquer outro plasmídeo bacteriano utilizado para clonagem de genes pode substituir o pBluescript KS+, na presente invenção. Dessa forma, a presente invenção não se limita ao uso do plasmídeo pBluescript KS+, para a obtenção dos clones provenientes do DNA-A e do DNA-B de SoCSV.

[37] Para obter o clone infeccioso, que corresponde a "cópia e meia" (1,5 cópia), os fragmentos correspondentes à cópia completa de cada componente do genoma viral foram inseridos nos respectivos clones "meia-cópia", que gerou a construção do clone infeccioso correspondente a 1,5 cópia do genoma, contendo duas origens de replicação na mesma orientação.

[38] O DNA plasmidial do clone pUFV1587, correspondente uma cópia do DNA-A SoCSV, foi digerido com a enzima de restrição *Xba*I, e recircularizado com a enzima T4 DNA ligase. O clone resultante, designado pUFV1589, contém um fragmento de aproximadamente 500 nucleotídeos que inclui a região comum e a origem de replicação que correspondeu ao clone "meia-cópia" do DNA-A. Uma cópia completa do DNA-A (2600pb) foi liberada do clone pUFV1587 pela clivagem com a enzima de restrição *Sac*I. O fragmento liberado foi purificado após separação em gel de agarose 1%, e inserido no clone pUFV1589 previamente digerido com a

mesma enzima *Sac*I, gerando o clone multimérico para o DNA-A, denominado clone infeccioso pUFV1591, correspondendo a 1,5 cópia do genoma, contendo as duas origens de replicação.

[39] Para a construção do clone infeccioso do DNA-B, o DNA plasmidial do clone pUFV1588, que contém uma cópia do DNA-B, foi submetido a clivagem com a enzima *Eco*RV e recircularizado produzindo o pUFV1590, que contém um fragmento de aproximadamente 800 pb, e corresponde à "meia-cópia" do DNA-B SoCSV. Uma cópia completa do DNA-B (2600pb) foi liberada do clone pUFV1588 pela clivagem com a enzima *Sac*I. O inserto foi purificado e inserido em pUFV1590 previamente digerido com a mesma enzima *Sac*I, gerando o clone multimérico para o DNA-B, denominado clone infeccioso pUFV1592 e identificado como **SEQ ID NO:4**. A 1,5 cópia do DNA-B de SoCSV, objeto do presente pedido de patente e presente no pUFV1592, é identificada como **SEQ ID NO:3**.

[40] A confirmação da ligação dos fragmentos de 2600 pb, tanto para DNA-A quanto para DNA-B em seus respectivos clones "meia-cópia", foi realizada por meio da clivagem das amostras de DNA plasmidial com as mesmas enzimas *Xba*I/*Sac*I, utilizadas na clonagem do clone infeccioso pUFV1591 (DNA-A SoCSV) (Figura 1A) e *Eco*RV/*Sac*I, utilizadas para obtenção do clone infeccioso pUFV1592 (DNA-B SoCSV) (Figura 1B) e por sequenciamento do DNA plasmidial.

[41] Para a construção de um segundo clone infeccioso para o DNA-B contendo uma cópia do DNA-B entre duas origens de replicação, o DNA plasmidial do clone pUFV2571 (que contém uma cópia completa do DNA-B) foi submetido à clivagem com a enzima *Xba*I e recircularizado produzindo o clone pUFV2572, que contém um fragmento de aproximadamente 600 pb, incluindo a região comum e a origem de replicação, e corresponde à "meia-cópia" do DNA-B SoCSV. Uma cópia completa do DNA-B (2600 pb) foi liberada do clone pUFV1588 pela clivagem com a enzima *Sac*I. O inserto foi purificado e inserido em pUFV2572, previamente digerido com a mesma enzima *Sac*I e desfosforilado, gerando o clone multimérico para o DNA-B, denominado clone infeccioso pUFV2573 e identificado como **SEQ ID NO:6**. A 1,5 cópia do DNA-B de SoCSV, objeto do presente pedido de patente e presente no pUFV2573, é identificada como **SEQ ID NO:5**.

[42] Os dois clones infecciosos de DNA-B, pUFV1592 e pUFV2573, foram confirmados por digestão com a enzima de restrição *Sma*I (Figura 2) e por sequenciamento do DNA plasmidial.

[43] A sequência parcialmente duplicada do genoma do SoCSV, clone infeccioso pUFV1591(DNA-A), clonada no vetor pBluescript KS+ pode ser utilizada juntamente com os clones infecciosos pUFV1592 (DNA-B) ou pUFV2573 (DNA-B) para inoculação e infeccção de plantas de soja. Em células inoculadas com as combinações de clones DNA-A e DNA-Bs infecciosos, os clones pUFV1591 e pUFV2573 liberam uma cópia do DNA-A e DNA-B, respectivamente, por replicação dos componentes virais, enquanto que o clone pUFV1592 libera uma cópia do DNA-B por recombinação do genoma viral.

[44] Um mapa de restrição foi gerado para a sequência do clone pUFV1591 utilizando o programa ApE-A Plasmid Editor. A enzima *Eco*NI que cliva somente na região que codifica a proteína do capsídeo na posição 1357 foi utilizada para analisar se a alteração de frame, ou seja, a interrupção da sequência aberta de leitura da proteína do capsídeo do vírus SoCSV (pUFV1591), não interfere na capacidade de infecção do vírus. Foi feita uma reação de digestão do clone pUFV1591 com *Eco*NI, em seguida foi utilizada DNA Polimerase de alta fidelidade para preenchimento das extensões 5' geradas da fita simples. Para a reação de preenchimento, foi utilizado 0,75 µl da reação de digestão, dNTPs 0,2 mM de cada nucleotídeo, 0.4 U de DNA Polimerase, 8 µl do tampão 5X (TAPS-HCl 125 mM, pH=9.3, KCl 250 mM, $MgCl_2$10 mM, β-mercaptoetanol 5 mM) totalizando um volume final de 40 µl. A reação foi conduzida em termociclador, à temperatura de 72°C por 10 min. Em seguida, o produto da reação foi recircularizado usando a enzima T4 DNA Ligase, gerando o clone pUFV1712 (clone denominado Frameshift) com interrupção da sequência aberta de leitura da proteína do capsídeo (CP). A confirmação dos clones com interrupção da sequência aberta de leitura da proteína CP foi simulada através de um padrão de digestão com as enzimas *Bam*HI e *Eco*NI por meio do uso do programa *ApE-A Plasmid Editor* e por reação de digestão utilizando as enzimas *Bam*HI e *Eco*NI, que apresentaram o padrão esperado observado na Figura 3.

[45] A infectividade do clone pUFV1712 (Frameshift) e pUFV1591 (DNA-A), ambos inoculados juntamente com pUFV1592 (DNA-B) ou pUFV2573 (DNA-B) foi comprovada por meio de um ensaio de inoculação utilizando o bombardeamento biolístico [Aragão F.J.L., Barros L.M.G., Brasileiro A.C.M., Ribeiro S.G., Smith F.D., Stanford J.C., Faria J.C., Rech, E.L. Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris L.*) co-transformed via particle bombardment. 1996. Theoretical and Applied Genetics. 93, 142-150]. A

dispensabilidade da CP para infecção viral em plantas de soja (Glycine max) foi confirmada, uma vez que as plantas inoculadas com o clone pUFV1712, juntamente com pUFV1592 ou pUFV2573, foram infectadas. Após 28 dias da inoculação, procedeu-se à extração de DNA pelo método do CTAB 2% das plantas de soja inoculadas e a infecção viral foi comprovada por meio de amplificação via PCR de um fragmento do genoma viral de 1200 pb, utilizando oligonucleotídeos específicos para a espécie SoCSV-DNA-A, sendo esses apresentados como as **SEQ ID NO:7** e **SEQ ID NO:8**. A infecção viral foi confirmada por PCR de DNA total de plantas inoculadas com os clones pUFV1591 + pUFV1592 (Figura 4, canaletas 5-8) e com os clones pUFV1712 + pUFV1592 (Figura 4, canaletas 9-12). A amplificação do fragmento do genoma viral de 1200 pb a partir do DNA total extraído de plantas inoculadas com o clone pUFV1712, com interrupção da sequência aberta de leitura da proteína do capsídeo, confirmou a dispensabilidade da CP para infecção viral. As reações de amplificação (PCR) continham: 50 ng de DNA, dNTPs 0,2 mM de cada nucleotídeo, 0,4 µM de cada oligonucleotídeo (SoCSVNcoI-R e SoCSVMCS-F), 0.2 U de Taq polimerase em um volume final de 50 µl. A reação de amplificação foi conduzida em termociclador, com uma etapa inicial de desnaturação a 94°C por 2 min, seguida por 35 ciclos (94°C por 45 s., 50°C por 45 s., e 68°C por 1 min e 45 s.) e um período adicional de polimerização a 68°C por 10 min. O produto de amplificação foi analisado por eletroforese em gel de agarose 1%, corado com brometo de etídio 0,1 µg/mL. A presença do fragmento amplificado de aproximadamente 1200 pb comprovou a infecção viral.

[46] A infecção viral também foi analisada e comprovada fenotipicamente quando comparada com a planta sem infecção (Figura 5). Sintomas suaves característicos da infecção viral, caracterizados por pontos cloróticos nas folhas de soja cultivar Conquista, foram observados tanto nas plantas inoculadas com o clone infeccioso pUFV1591 (DNA-A) + pUFV1592 (DNA B) como naquelas inoculadas com os clones pUFV1712 (Frameshift) + pUFV1592 (DNA B), respectivamente (Figuras 5A.2 e 5A.3), em relação a planta controle sem inoculação (Figura 5A.1), confirmando a dispensabilidade da CP para infecção viral. O mesmo resultado foi verificado e comprovado pela inoculação dos clones pUFV1591 (DNA-A) e pUFV1712 (Frameshift) juntamente com pUFV2573 (DNA B) observado nas figuras 5B.2 e 5B.3, respectivamente, em relação a planta controle sem inoculação (Figura 5B.1).

[47] Uma vez confirmada que a CP não é requerida para a infecção viral sistêmica, procedeu-se uma análise do padrão de restrição do clone pUFV1591 utilizando-se o programa *ApE-A Plasmid Editor* e a enzima NcoI foi selecionada para retirar a região do gene que codifica a proteína do capsídeo (CP). A enzima NcoI cliva nas sequências do inserto do clone pUFV1591 (nas posições 1303 e 3029) e, após digestão, libera um fragmento de aproximadamente 1726 pb. O clone pUFV1591 foi clivado com a enzima NcoI, e o produto da digestão foi analisado por eletroforese em gel de agarose 1%, corado com brometo de etídio 0,1 µg/mL, conforme padrão verificado na Figura 6, canaletas 1-5. O fragmento correspondente a aproximadamente 4385 pb (pUFV1591 menos o fragmento *Nco*I de 1726 pb) foi purificado do gel de agarose.

[48] Durante a reação de clivagem do clone pUFV1591 com *Nco*I, foi liberada uma grande parte da região que codifica a CP e também foi retirada a região que codifica as proteínas Ren e Trap e uma parte da sequência que codifica a proteína Rep, que são essenciais para o vírus. Para a reconstituição desses genes, foram desenhados oligonucleotídeos para amplificação das regiões perdidas e para a inserção de um sítio múltiplo de clonagem, sendo tais oligonucleotídeos as **SEQ ID NO:7** e **SEQ ID NO:8**. Para tal, a reação de amplificação conteve: 50 ng de DNA plasmidial do clone pUFV1591, dNTPs 0,2 mM de cada dNTP, 0,4 µM de cada oligonucleotídeo (SoCSVNcoI-R e SoCSVMCS-F), 0.2 U de Taq Polimerase de alta fidelidade em um volume final de 50 µl. A reação de amplificação foi conduzida no termociclador, com uma etapa inicial de desnaturação a 94°C por 2 min., seguida por 35 ciclos (94°C por 45 s., 50°C por 45 s., e 68°C por 1 min e 45 s.) e um período adicional de polimerização a 68°C por 10 min. O produto de amplificação foi analisado por eletroforese em gel de agarose 1%, corado com brometo de etídio 0,1 µg/mL (Figura 6 – canaleta 6). O fragmento amplificado de aproximadamente 1200 pares de base (pb) foi purificado do gel de agarose.

[49] O produto da amplificação de 1200 pb purificado foi ligado utilizando-se a enzima T4 DNA Ligase, ao fragmento de 4385 pb do clone infeccioso pUFV1591. O produto da reação de ligação foi utilizado para transformação de Escherichia coli DH5α pelo método de choque térmico (Sambrook J., Fritsch E.F. Maniatis T. Molecular Cloning – A Laboratory Manual. 2001. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press) e as colônias contendo os plasmídeos recombinantes foram submetidas a minipreparação de DNA plasmidial. O DNA

plasmidial dos possíveis recombinantes foi digerido com NcoI e confirmado conforme padrão presente na canaleta 3 da Figura 7. Para confirmar se a orientação do inserto estava correta, foi feito a simulação de um padrão de restrição com as enzimas XbaI e NdeI usando o programa *ApE-A Plasmid Editor* (Figura 8). Na Figura 9, verifica-se o padrão de restrição esperado da colônia confirmada clivada com as enzimas XbaI e NdeI do DNA plasmidial do clone pUFV1591 usado como controle (canaleta 1 da Figura 9) e do plasmídeo recombinante para obtenção do vetor para silenciamento gênico induzido por vírus (canaleta 2 da Figura 9). O clone resultante, pUFV1713, identificado também como SEQ ID NO:2, contém a sequência parcialmente duplicada do DNA-A do isolado SoCSV (pUFV1591), sem a maior parte da região codificante da CP e com um sítio múltiplo de clonagem para as enzimas BglII, HpaI e NheI, apresentada como SEQ ID NO:1, correpondente, portanto, ao vetor de silenciamento gênico induzido por vírus, Figura 10. A clivagem com as enzimas BglII e NheI gera extremidades coesivas e a clivagem com a enzima *Hpa*I gera extremidades abruptas, o que pode facilitar a clonagem, caso se deseje inserir um fragmento que não possua sítios de clivagem para as enzimas presentes no vetor.

[50] A Figura 10 é o mapa do vetor para VIGS construído, indicando o sítio múltiplo de clonagem nas extremidades do gene truncado da CP. A direção e extensão das ORFs virais são indicadas pelas setas e o *stem loop* está destacado em duplicata (seta em vermelho), mostrando que as sequências virais foram clonadas entre duas origens de replicação do vírus. Consequentemente, após a inoculação via biobalística nas células vegetais, um episssoma circular é liberado do plasmídeo recombinante por replicação e este replica independentemente do genoma do hospedeiro. A infecção viral foi confirmada via PCR, utilizando oligonucleotídeos específicos, que delimitam uma banda correspondente ao tamanho de 1200 pb visualizado em gel de agarose, como mostrado na Figura 4. Desta forma, este vetor de silenciamento gênico viral pode ser utilizado para estudos de genes funcionais em plantas de soja.

[51] Os clones pUFV1592 e pUFV2573 correspondem aos clones infecciosos para o componente B de SoCSV e foram usados (um ou outro) com a construção do vetor de silenciamento viral para as inoculações aqui descritas, a menos que indicado de outra forma. Os dois clones para o componente B são eficientes em causar a infecção viral na presença do componente A, o que pode ser observado na Figura 5

pela presença de sintomas tanto em plantas inoculadas com pUFV1591 (DNA-A SoCSV) + pUFV1592 ou pUFV1591 (DNA-A SoCSV) + pUFV2573, assim como pUFV1713 (VIGS) + pUFV1592 ou pUFV1713 (VIGS) + pUFV2573.

EXPERIMENTO DE DEMONSTRAÇÃO

Clonagem do fragmento do gene *ChlI (magnesium chelatase subunit I)* no vetor de silenciamento gênico viral derivado da modificação do genoma do SoCSV (pUFV1713)

[52] Para a clonagem do fragmento do gene *ChlI (magnesium chelatase subunit I),* enzima envolvida na formação de clorofila, no vetor de silenciamento gênico viral derivado da modificação do genoma do vírus SoCSV (pUFV1713), primeiramente foi feito a extração de RNA da planta de soja, seguido da síntese do cDNA para isolamento do fragmento do gene *Chll* com oligonucleotídeos específicos previamente desenhados e posterior clonagem deste fragmento no vetor pUFV1713 para VIGS.

[53] O RNA total das folhas de soja foi extraído utilizando o reagente Trizol. Para eliminação do DNA contaminante, o RNA total foi tratado com uma unidade de DNase livre de RNase. O RNA foi quantificado e analisado em gel de agarose desnaturante 1,5% (p/v), corado com brometo de etídio 0,1 µg/mL. Posteriormente, a síntese de cDNA foi realizada utilizando 4 µg de RNA total, oligo-dT (18) e a enzima Trancriptase Reversa.

[54] O fragmento do gene *ChlI*, correspondente a 350 pb, foi isolado via PCR, utilizando o cDNA sintetizado a partir do RNA total de soja. A reação de amplificação conteve: 1 µl da reação de síntese de cDNA, dNTPs 0,2 mM de cada nucleotídeo, 0,4 µM de cada oligonucleotídeo, **SEQ ID NO:9** e **SEQ ID NO:10**, e 0.2 U de Taq polimerase de alta fidelidade em um volume final de 50 µl.

[55] A reação de amplificação foi conduzida em termociclador, com uma etapa inicial de desnaturação a 94°C por 3 min., seguida por 30 ciclos (94°C por 45 s., 57°C por 45 s. e 68°C por 45 s.) e um período adicional de polimerização a 68°C por 10 min. O produto de amplificação foi analisado por eletroforese em gel de agarose 1%, corado com brometo de etídio 0,1 µg/mL. O padrão de bandas do DNA foi visualizado sob luz ultravioleta e fotodocumentado.

[56] Após a confirmação da amplificação, o fragmento obtido do gene *Chll* foi purificado, utilizando-se um Kit de extração de fragmentos de DNA de géis de agarose, e clonado no vetor de entrada pCR8/GW/TOPO. O produto da reação foi utilizado para transformação de E. coli DH5α pelo método de choque térmico. O fragmento do gene *ChlI* foi liberado do vetor de entrada comercial pCR8/GW/TOPO por clivagem com a enzima de restrição *Eco*RI e as extensões 5' fita simples foram preenchidas utilizando-se a enzima DNA Polimerase I. Em seguida, o fragmento do gene *ChlI* foi inserido no vetor viral de silenciamento (pUFV1713), previamente clivado no seu sítio múltiplo de clonagem com a enzima de restrição *Hpa*I e desfosforilado, utilizando-se a enzima T4 DNA Ligase. O produto da reação de ligação foi utilizado para transformação de E. Coli DH5α pelo método de choque térmico e as colônias contendo os plasmídeos recombinantes foram submetidas a minipreparação de DNA plasmidial.

[57] Na Figura 11, verifica-se o perfil das bandas de confirmação dos DNA recombinantes como resultado da clivagem com as enzimas BglII e NheI. O clone positivo resultante foi denominado pUFV1732.

[58] O clone pUFV1732 foi confirmado por sequenciamento e corresponde ao vetor de silenciamento gênico viral contendo um fragmento de 350 pb do gene ChII (VIGS-Chll). Este clone foi utilizado para silencimento do gene ChII de plantas de soja e comprovação da efetividade do silenciamento induzido pelo vertor construído para VIGS.

Efetividade do vetor construído para VIGS por meio da indução do silenciamento do gene *ChlI (magnesium chelatase subunit I)* em plantas de soja inoculadas com o vetor pUFV1732 contendo um fragmento de 350 pb do gene ChII (VIGS-Chll) e o DNA-B SoCSV.

[59] Com o objetivo de comprovar a efetividade do vetor desenvolvido para VIGS, induziu-se o silenciamento do gene *Chll* de soja utilizando o vetor desenvolvido (pUFV1732, VIGS-Chll) e procedeu-se à comprovação por meio dos métodos:

a. confirmação da infecção por PCR (Figura 4)
b. avaliação do fenótipo (Figura 5)

c. quantificação por *real time* dos níveis de expressão do mRNA do gene *ChlI* em comparação com os níveis de expressão de uma planta infectada com vetor VIGS vazio (Figura 12)

d. análise de *northern blot* (Figura 13) para detecção de possíveis siRNA do tamanho esperado de 21 nt gerados pelo processo de silenciamento gênico.

[60] Para realização deste experimento de demonstração, as sementes de soja do cultivar Conquista foram inicialmente germinadas em um germinador regulado a 25ºC, utilizando papel do tipo "Germitest" como substrato para semeadura, na forma de rolo, umedecido com água em 2,5 vezes o peso do papel seco. Após três dias, com a protrusão radicular, as radículas foram inoculadas via biobalística, com o DNA plasmidial das seguintes combinações de clones:

a. clone pUFV1732, que corresponde ao vetor de silenciamento gênico induzido por vírus contendo um fragmento de 350 pb do gene Chll (VIGS-Chll) + pUFV1592 (DNA-B SoCSV)

b. pUFV1732 (VIGS-Chll) + pUFV2573 (DNA-B SoCSV)

c. pUFV1713 (VIGS), utilizado como controle positivo da inoculação + pUFV1592 (DNA-B SoCSV)

d. pUFV1713 + pUFV2573 (DNA-B SoCSV)

e. controle negativo que corresponde à inoculação com partículas de tungstênio sem DNA viral.

[61] A infecção viral foi confirmada via PCR, 28 dias pós-inoculação. Na Figura 4, verifica-se a confirmação da infecção viral das plantas de soja por PCR, após 28 dias da inoculação, utilizando-se os oligonucleotídeos específicos SoCSVNcoI-Rvs e SoCSVMCS-Fwd para o DNA-A. A banda específica de tamanho esperado de 1200 pb, visualizado em gel de agarose 1%, foi amplificada de DNA total extraído de plantas inoculadas com as construções pUFV1713 (VIGS) + pUFV1592 (DNA B) nas canaletas 13-16 pUFV1732 (VIGS-Chll) + pUFV1592 (DNA B) na canaleta 17 e pUFV1732 (VIGS-Chll) + pUFV2573 (DNA B) nas canaletas 18 e 19.

[62] Os sintomas visuais de silenciamento foram verificados a partir de 35 dias após inoculação de plantas com a construção pUFV1732 (VIGS-Chll) juntamente com DNA B (pUFV1592). Observa-se que as plantas inoculadas com a construção pUFV1732 (VIGS-Chll) juntamente com DNA B (pUFV1592) possuem amarelecimento intenso das folhas, sintoma característico do silenciamento do gene

Chll, como observado na Figura 5A.5. Estes sintomas são mais severos do que o causado pela infecção viral resultante da inoculação de plantas com o clone infeccioso pUFV1591 + pUFV1592 (Figura 5A.2), o clone com interrupção da sequência da proteína da capa protéica, pUFV1713 + pUFV1592 (Figura 5A.3) ou vetor vazio para VIGS (puFV1713) + pUFV1592 (Figura 5A.4). Constatou-se assim um amarelecimento mais intenso das folhas em plantas inoculadas com o vetor pUFV1732 (VIGS-Chll) + pUFV1592, do que o fenótipo provocado pela infecção por SoCSV (pUFV1591 + pUFV1592), caracterizada por pontos cloróticos, comprovando o silenciamento gênico de *Chll* em plantas inoculadas com o vetor pUFV1732 (VIGS-Chll).

[63] Os sintomas visuais de silenciamento também foram verificados a partir de 35 dias após inoculação da planta pelo clone pUFV1732 (VIGS-Chll) juntamente com DNA B (pUFV2573). Foi observado que as plantas silenciadas possuem amarelecimento das folhas, sintoma característico do silenciamento do gene *Chll*, como observado nas Figuras 5B.5 e 5C.5. Estes sintomas são mais severos que o causado pela infecção viral quando comparado ao inocular a planta com os clones infecciosos pUFV1591 + pUFV2573 (Figuras 5B.2 e 5C.2), o clone com interrupção da sequência da proteína da capa protéica, pUFV1712 + pUFV2573 (Figuras 5B.3 e 5C.3) ou o vetor vazio para VIGS (pUFV1713) + pUFV2573 (Figuras 5B.4 e 5C.4). Foi constatado que um amarelecimento mais intenso das folhas em plantas inoculadas com o vetor pUFV1732 (VIGS-Chll) + pUFV 2573, do que aquele provocado pela inoculação com pUFV1732 (VIGS-Chll) + pUFV1592 (compare Figuras 5A.5 e 5B.5), indicando uma possível maior eficiência de replicação da cópia e meia de DNA-B por meio do clone pUFV2573.

[64] Para realização do *real time* e *northern blot*, o RNA total de folhas de soja das plantas infectadas após 28 dias, 35 e 42 dias após inoculação foi extraído utilizando o reagente Trizol, seguindo as recomendações do fabricante. Para eliminação do DNA contaminante, o RNA total foi tratado com uma unidade de DNase livre de RNase. O RNA foi quantificado por A_{260nm} e analisado em gel de agarose desnaturante 1,5 % (p/v), corado com brometo de etídio 0,1 µg/mL. A síntese de cDNA foi realizada utilizando 4 µg de RNA total, oligo-dT(18) e Trancriptase Reversa. O cDNA sintetizado foi utilizado na avaliação do decaimento do gene ChlI por PCR em tempo real (Figura 12) e o RNA total foi utilizado na análise de *Northen blot* para detecção siRNAs (Figura 13).

[65] Para avaliar os níveis de expressão do gene *ChlI*, foram utilizados os oligonucleotídeos específicos, **SEQ ID NO:11** e **SEQ ID NO:12**, cDNAs dos tratamentos e SYBR Green PCR Master Mix. As condições de amplificação foram: 50°C por 2 minutos, 95°C por 10 minutos, e 40 ciclos de 94°C por 15 segundos e 60°C por 1 minuto. Para a quantificação da expressão gênica, foi utilizado o método comparativo de Ct: $2^{-\Delta Ct}$. Como controle endógeno para normalização dos dados do qRT-PCR, foi utilizado o oligonucleotídeo específico para o gene de soja 18S, **SEQ ID NO:13** e **SEQ ID NO:14**, validado como gene normalizador pelo programa geNorm (http://medgen.ugent.be/~jvdesomp/genorm/).

[66] Os siRNAs gerados provenientes do silenciamento gênico foram detectados por *northern blot*. Para o procedimento de *northen blot*, o RNA total de folhas de soja das plantas infectadas após 28 dias, 35 e 42 dias após inoculação com vetor vazio para VIGS (pUFV1713) e o vetor pUFV1732, contendo um fragmento de 350 pb do gene ChII (VIGS-Chll) foi extraído, utilizando o reagente Trizol. 120 µg do RNA total foram separadas por gel de acrilamida 15% contendo 7 M de ureia, Tris-borato-EDTA (TBE) 1x e adição de 10% persulfato de amônia e TEMED. Quatro oligonucleotídeos (18, 24, 39 e 44 nucleotídeos) foram utilizados como marcadores de tamanho molecular. Após aproximadamente 3 horas de corrida a 210 V, o gel foi corado em solução de TBE 1X contendo brometo de etídio e fotografado para verificar o padrão do RNA. Em seguida, o RNA foi transferido por 4 horas a 0,20 A, para uma membrana IMMOBILON-NY + de Nylon e fixados na luz UV. A hibridização dos RNAs foi realizada a 45ºC, overnight, com uma sonda de 100 ng de DNA correspondente ao fragmento de 149 pb do gene *Chll*, isolado previamente por PCR, utilizando cDNA sintetizado a partir do RNA total de soja. A reação de amplificação da sonda continha: 1 µl da reação de síntese de cDNA, dNTPs 0,2 mM de cada nucleotídeo, 0,4 µM de cada oligonucleotídeo, **SEQ ID NO:15** e **SEQ ID NO:16**, e 0.2 U de Taq polimerase de alta fidelidade em um volume final de 50 µl. A reação de amplificação do fragmento a ser usado como sonda foi conduzida no termociclador, com uma etapa inicial de desnaturação a 94 °C por 3 min., seguida por 30 ciclos (94°C por 45 s., 57°C por 45 s., e 68°C por 45 s.) e um período adicional de polimerização a 68°C por 10 min. O produto de amplificação da sonda foi analisado por eletroforese em gel de agarose 1%, corado com brometo de etídio 0,1 µg/mL e o padrão de bandas do DNA foi visualizado sob luz ultravioleta e fotodocumentado. Após a confirmação da amplificação, o fragmento obtido foi

purificado utilizando-se kit de extração. As sondas foram marcadas com $^{\alpha\text{-}32}$P dCTP usando Thermo Scientific DecaLabel DNA Labeling e autoradiografadas.

[67] O silenciamento do gene *Chll* também foi avaliado nas plantas infectadas por PCR em tempo real (Figura 12). As plantas infectadas com pUFV1732 (VIGS-ChlI) + pUFV1592 ou pUFV2573 apresentaram diminuição dos níveis de expressão do mRNA do gene ChlI em comparação com os níveis de expressão de uma planta infectada com vetor vazio para VIGS (pUFV1713), após 42 dias de inoculação (Figure 12). O nível de decaimento dos transcritos de ChII em plantas de soja inoculadas não foi relacionado com o acúmulo de siRNA de 21 nucleotídeos em plantas infectadas que foram inoculadas com construção pUFV1732 (VIGS-Chll Figura 13). Contudo, o vetor pUFV1732 induziu níveis diferentes de silenciamento do gene *ChlI* em plantas inoculadas. RNAs de 21 bases (siRNA) foram detectados em plantas de soja inoculadas com o DNA plasmidial pUFV1732 (VIGS contendo o fragmento do gene *ChlI*) + pUFV1592 (DNA-B de SoCSV), correspondente à canaleta 7 da Figura 13 e com a o DNA plasmidial pUFV1732 (VIGS contendo o fragmento do gene *ChlI*) + pUFV2573 (DNA-B de SoCSV), canaleta 8 da Figura 13.

[68] Em resumo, o vetor de silenciamento pUFV1732 contendo o fragmento do gene *ChlI (*VIGS-Chll) promoveu o silenciamento do gene alvo, o que foi comprovada fenotipicamente através do amarelecimento das folhas de soja (Figura 5A.5 e 5B.5.), um fenótipo típico do silenciamento do gene ChlI (Jeddeloh J.A., Bender J., Richards E.J. The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis. 1998. Genes Dev., 12:1714–25), por meio do decaimento de transcritos do gene ChlI monitorado por PCR em tempo real (Figura 12) e por meio da detecção de siRNAs de sequencias do gene *Chll* por *northen blot* (Figura 13, canaletas 7 e 8). Coletivamente, estes resultados demonstraram que o vetor de silenciamento gênico desenvolvido, baseado na modificação do genoma de SoCSV, é eficiente em silenciar genes endógenos, como *ChII*, da planta hospedeira soja e, consequentemente, funciona como um vetor para VIGS.

[69] Os resultados demonstram que o vetor construído pode silenciar ativamente a expressão do gene endógeno e pode proporcionar um mecanismo para a supressão sistêmica de expressão de genes em plantas. A presente invenção permite a modulação da expressão de genes em plantas, sem a necessidade de transformação estável de plantas. Mais especificamente, é demonstrado a habilidade de silenciar a

expressão de genes endógenos sistemicamente em uma planta usando uma construção de um vírus de replicação episssomal no núcleo de células infectadas.

CONCLUSÃO

[70] O vetor para VIGS desenvolvido, objeto do presente pedido de patente, é uma ferramenta de extrema importância para análise funcional de genes em leguminosas. O desenvolvimento do vetor de silenciamento viral baseado na modificação no genoma do vírus *Soybean chlorotic spot virus* (SoCSV) para induzir o silenciamento sistêmico de genes endógenos de plantas pertencentes à família *Fabaceae*, especialmente soja, oferece grandes vantagens por ser mais rápida e permitir a análise de genes em larga escala, sendo extremamente útil no estudo da expressão de genes de soja. Esta ferramenta molecular desenvolvida permite um avanço no conhecimento básico de funções gênicas desconhecidas e na consequente seleção de genes de interesse para o melhoramento genético da soja, cultura de elevada importância econômica para o Brasil.

REIVINDICAÇÕES

1) VETOR VIRAL caracterizado por compreender os seguintes componentes:

a) SEQ ID NO:1 inserida em um plasmídeo bacteriano e

b) SEQ ID NO:3 ou SEQ ID NO:5 inserida em um plasmídeo bacteriano.

2) VETOR VIRAL, de acordo com a reivindicação 1, caracterizado pelo plasmídeo bacteriano ser qualquer plasmídeo bacteriano apto para clonagem gênica.

3) VETOR VIRAL, de acordo com a reivindicação 1, caracterizado pela SEQ ID NO:1 compreender o genoma modificado do DNA-A de *Soybean Chlorotic spot virus* (SoCSV), possuindo na direção 5' para 3':

- Uma origem de replicação de SoCSV
- Uma sequência truncada que codifica a região amino terminal da proteína capsidial
- Um sítio múltiplo de clonagem
- Uma sequência que codifica para as proteínas virais Ren, TrAP, Rep e AC4
- Uma origem de replicação repetida.

4) VETOR VIRAL, de acordo com as reivindicações 1 e 3, caracterizado pela sequência da SEQ ID NO:1 que codifica para a proteína capsidial do SoCSV ser parcialmente substituída por um sítio múltiplo de clonagem contendo sítio de reconhecimento das enzimas de restrição *Bgl*II, *Hpa*I e *Nhe*I.

5) VETOR VIRAL, de acordo com as reivindicações 1, 3, 4 e 5, caracterizado pela SEQ ID NO:1 ser capaz de conter um DNA heterólogo, em que este DNA heterólogo é idêntico ou com grande semelhança a sequências de um gene endógeno de plantas da família *Fabaceae.*

6) VETOR VIRAL, de acordo com a reivindicação 1, caracterizado pela SEQ ID NO:3 compreender o genoma modificado do DNA-B de *Soybean Chlorotic spot virus* (SoCSV), possuindo na direção 5' para 3':

- Uma origem de replicação de SoCSV
- Uma sequência que codifica para as proteínas virais NSP e MP
- Uma origem de replicação repetida.

7) VETOR VIRAL, de acordo com a reivindicação 1, caracterizado pela SEQ ID NO:5 compreender o genoma modificado do DNA-B de *Soybean Chlorotic spot virus* (SoCSV), possuindo, na direção 5’ para 3':

- Uma origem de replicação de SoCSV
- Uma sequência que codifica para a proteínas virais NSP e MP
- Uma sequência repetida de SoCSV da posição +1 à posição +771 do DNA-B do genoma de SoCSV.

8) USO DO VETOR VIRAL, definido na reivindicação 1, caracterizado por ser para o silenciamento de genes de plantas da família *Fabaceae***.**

FIGURAS E DESCRIÇÃO DAS FIGURAS

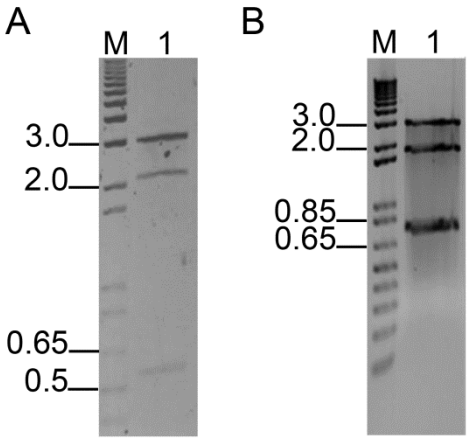


Figura 1: Padrão eletroforético em gel de agarose (1%) dos clones infecciosos, A- pUFV1591 (DNA-A SoCSV) e B- pUFV1592 (DNA-B SoCSV). Em A e B, a letra M refere-se ao marcador de tamanho de fragmentos de DNA em kpb. Em A, a canaleta 1 corresponde ao padrão da clivagem com as enzimas *Sac*I e *Xba*I do DNA plasmidial do clone infeccioso pUFV1591. Em B, e canaleta 2, ao padrão da clivagem com as enzimas *Sac*I e *Eco*RV do DNA plasmidial do clone infeccioso pUFV1592.

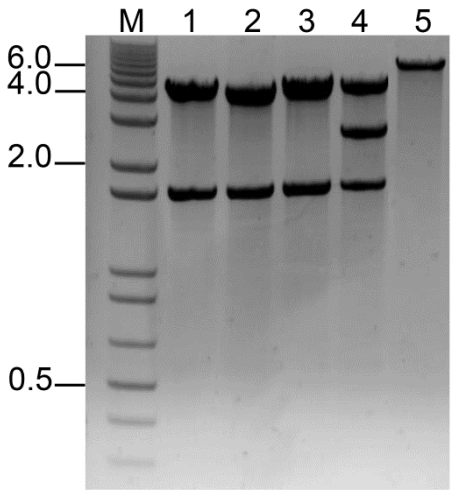


Figura 2: Padrão eletroforético em gel de agarose (1%) dos clones infecciosos pUFV2573 e pUFV1592 para DNA-B de SoCSV. A letra M refere-se ao marcador de tamanho de DNA em kpb. Nas canaletas 1 a 3, verifica-se o padrão da clivagem com a enzimas *SmaI* do DNA plasmidial do clone infeccioso pUFV2573, na canaleta 4, verifica-se o padrão de um clone não confirmado, na canaleta 5, verifica-se o padrão da clivagem com a enzimas *SmaI* do DNA plasmidial do clone infeccioso pUFV1592.

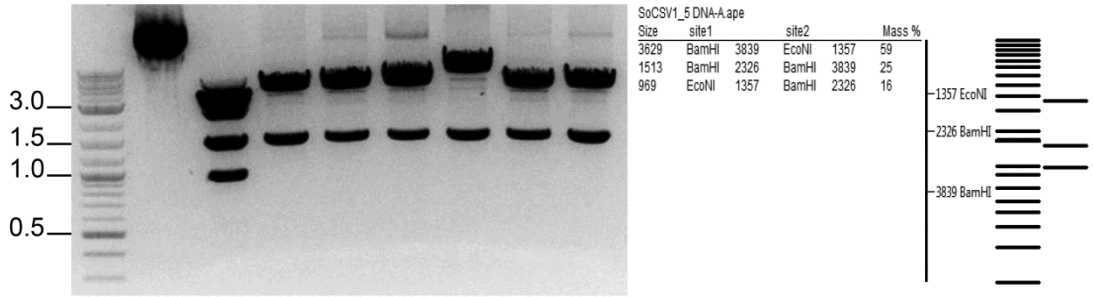


Figura 3: Padrão eletroforético em gel de agarose (1%) para diagnóstico dos possíveis clones contendo a interrupção da sequência aberta de leitura da proteína CP (pUFV1712). A letra M refere-se ao marcador de tamanho de DNA em kpb (“1Kb plus DNA ladder”) A canaleta 1 corresponde ao controle, DNA plasmidial do clone pUFV1591 canaleta 2, clivagem de pUFV1591 com as enzimas *Bam*HI e *Eco*NI 3-8, clivagem com as enzimas *Bam*HI e *Eco*NI dos possíveis clones contendo a interrupção da sequência aberta de leitura da proteína CP (pUFV1712). O lado direito da Figura mostra uma simulação do padrão de restrição esperado da digestão com *Bam*HI e *Eco*NI do clone pUFV1591, utilizando o programa APE-A Plasmid Editor.

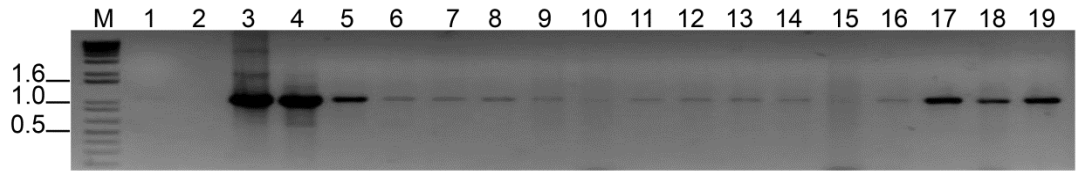


Figura 4: Diagnóstico molecular, via PCR, da infecção viral das plantas inoculadas com as construções de DNA dos clones infecciosos de DNA-A e DNA-B. A infecção é confirmada pelo aparecimento da banda de aproximadamente 1200 pb específica para SoCSV-DNA-A, após 28 dias de inoculação com as construções de DNA pUFV1591 (DNA-A SoCSV) em combinação com pUFV1592 (DNA-B SoCSV) ou pUFV2573 (DNA-B SoCSV). Plantas também foram inoculadas com partículas de tungstênio sem DNA viral, para controle negativo. Após 28 dias de inoculação, o DNA total foi extraído de folhas sistêmicas de plantas inoculadas com as construções acima citadas utilizadas como molde em reações de PCR. A letra M refere-se ao marcador de tamanho de DNA em kpb ("1Kb plus DNA ladder"). Os números 1-2 correspondem ao controle negativo, resultado do PCR do DNA total de plantas inoculadas com partículas de tungstênio sem DNA viral. Os números 3-4, controle positivo, resultado do PCR a partir do DNA plasmidial do clone pUFV1591. Os números 5-8, resultado do PCR do DNA total de plantas inoculadas com partículas de tungstênio contendo o DNA plasmidial do clone pUFV1591 + pUFV1592 . Os números 9-12, resultado do PCR do DNA total de plantas inoculadas com partículas de tungstênio contendo o DNA plasmidial do clone pUFV1712 (Frameshift) + pUFV1592. Os números 13-16, resultado do PCR do DNA total de plantas inoculadas com partículas de tungstênio contendo o DNA plasmidial do clone pUFV1713 (VIGS-vetor vazio) + pUFV1592. O números 17, resultado do PCR do DNA total de plantas inoculadas com partículas de tungstênio contendo o DNA plasmidial do clone pUFV1732 (VIGS-Chll) + pUFV1592 e números 18-19, resultado do PCR do DNA total de plantas inoculadas com partículas de tungstênio, contendo o DNA plasmidial do clone pUFV1732 (VIGS-Chll) + pUFV2573.

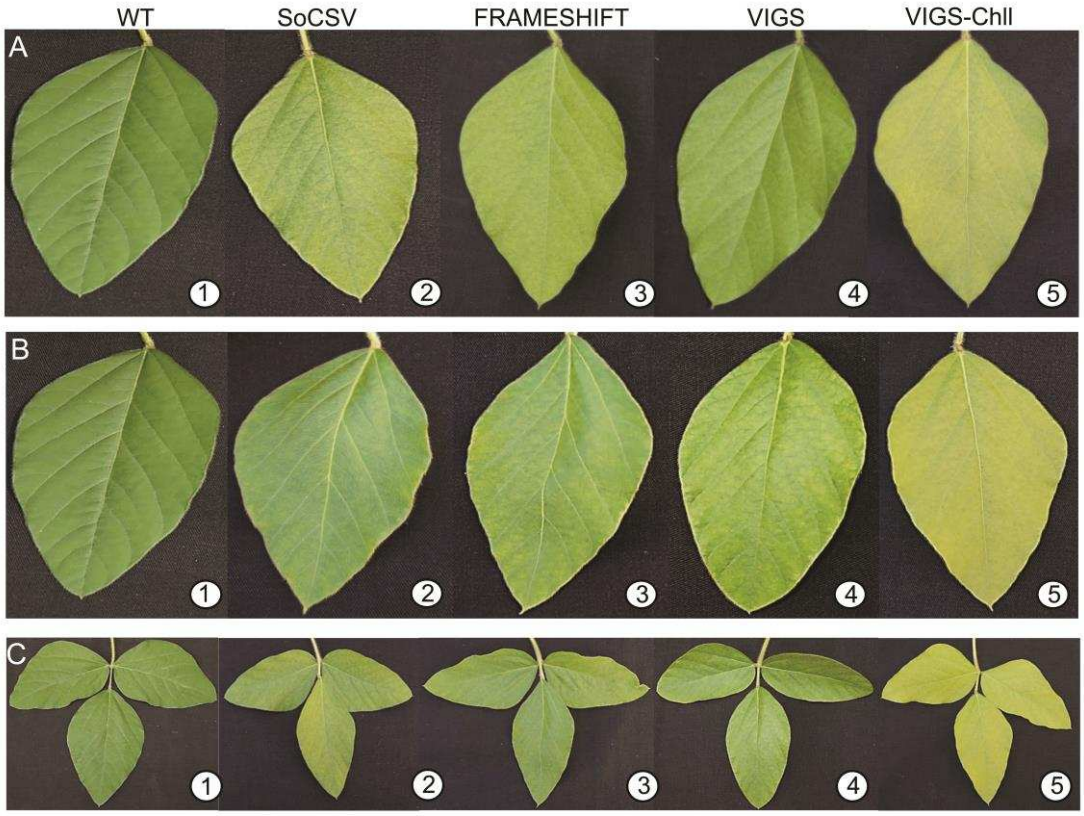


Figura 5: Sintomas de infecção viral e de silenciamento do gene Chll. A. Sintomas em folhas de soja (*Glycine max*), cultivar conquista, inoculadas via biobalística com combinações dos clones infecciosos de DNA-A derivados de pUFV1591 + o clone infeccioso DNA-B, pUFV1592. WT representa o controle negativo, correspondente à planta WT sem sintomas de infecção. SoCSV representa folhas de plantas inoculadas com partículas de tungstênio contendo o DNA plasmidial do clone infeccioso pUFV1591 (SoCSV) + pUFV1592 (DNA-B). FRAMESHIFT indica a inoculação com pUFV1712 (Frameshift) + pUFV1592 (DNA-B). VIGS, inoculação com pUFV1713 (VIGS-vetor vazio) + pUFV1592 (DNA-B). VIGS-Chll, inoculação com pUFV1732 (VIGS-Chll) + pUFV1592 (DNA-B), com sintomas do silenciamento gênico do gene ChlI *(magnesium chelatase subunit I)*. B. Folhas de soja (*Glycine max*) cultivar conquista inoculadas via biobalística com partículas de tungstênio sem DNA viral, para controle negativo, correspondente à planta WT sem sintomas de infecção, e com partículas de tungstênio contendo DNA B (pUFV2573) e as seguintes combinações de clones: DNA plasmidial do clone infeccioso pUFV1591 (SoCSV), pUFV1712 (Frameshift), pUFV1713 (VIGS-vetor vazio) ou pUFV1732 (VIGS-Chll). Note os sintomas do silenciamento gênico do gene ChlI *(magnesium chelatase subunit I)* em VIGS-ChlI. C. Sintomas exibidos pelos trifólios do cultivar Conquista, de plantas inoculadas com as mesmas construções descritas em B.

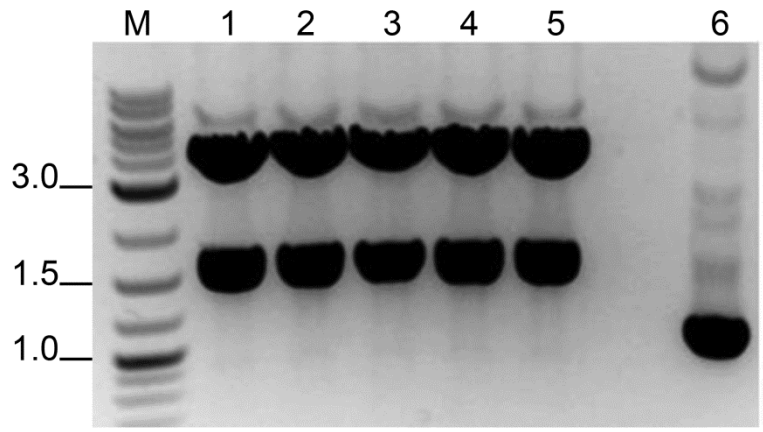


Figura 6: Padrão eletroforético em gel de agarose (1%) de clones deletados para região que codifica a proteína da capa protéica (CP). A letra M corresponde ao marcador de tamanho de DNA em kpb ("1Kb plus DNA ladder"). As canaletas 1-5 correspondem ao padrão de clivagem do clone pUFV1591 com a enzima *Nco*I. A canaleta 6 mostra o resultado do PCR a partir do clone pUFV1591 para amplificação de sequências importantes do genoma e inserção de um sítio múltiplo de clonagem.

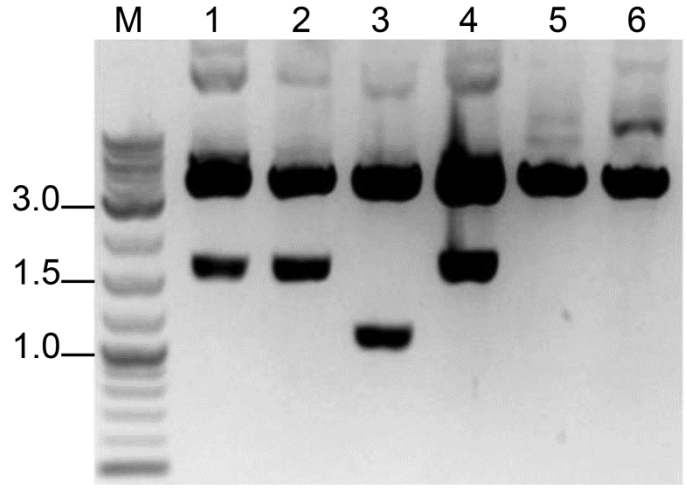


Figura 7: Padrão eletroforético em gel de agarose (1%) das minipreparações de DNA plasmidial clivadas com *Nco*I obtidas a partir de colônias contendo possíveis plasmídeos recombinates para obtenção do vetor de silencimento gênico induzido por vírus. A letra M corresponde ao marcador de tamanho de DNA em kpb ("1Kb plus DNA ladder") A canaleta 1, controle, DNA plasmidial do clone pUFV1591 clivado com *Nco*I. As canaletas 2-6, clivagem com *Nco*I das minipreparações de DNA plasmidial contendo os possíveis plasmídeos recombinates (pUFV1713).

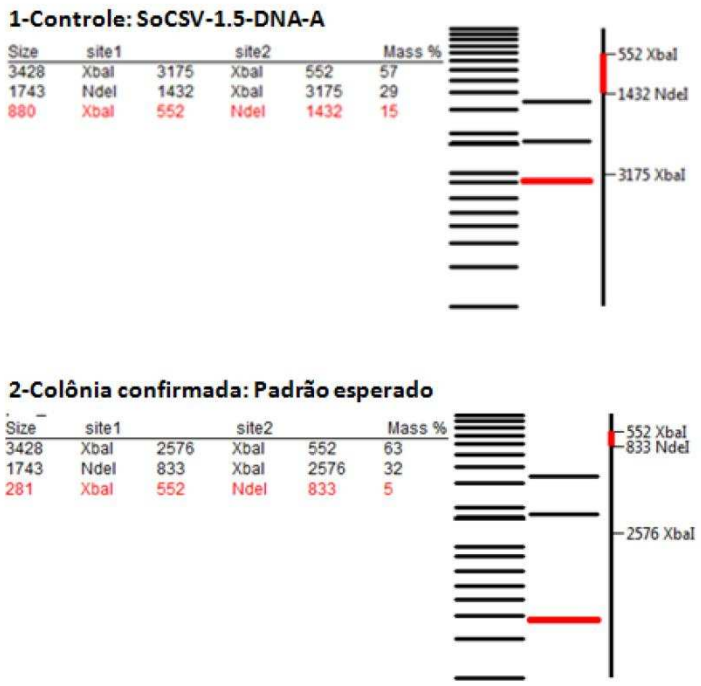


Figura 8: 8.1 é resultado da simulação do padrão de restrição esperado pela digestão com *Xba*I e *Nde*I do clone pUFV1591, utilizando o programa ApE-A Plasmid Editor. 8.2 representa a simulação do padrão de restrição esperado da digestão com *Xba*I e *Nde*I do plasmídeo recombinante esperado para obtenção do vetor de silencimento gênico induzido por vírus.

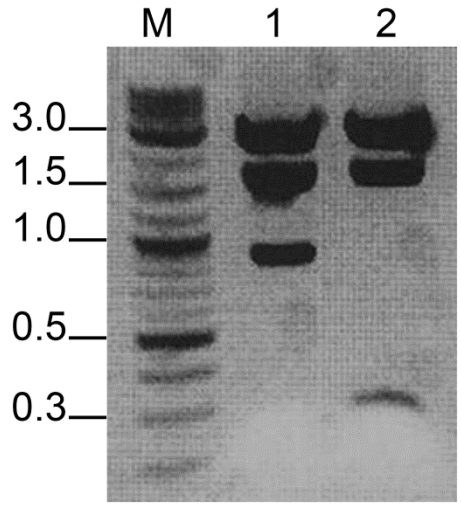


Figura 9: Padrão eletroforético em gel de agarose (1%) para diagnóstico do vetor de silenciamento gênico induzido por vírus (pUFV1713). A letra M corresponde ao marcador de tamanho de DNA em kpb (“1Kb plus DNA ladder”). A canaleta 1, controle que corresponde ao padrão de clivagem com as enzimas *Xba*I e *Nde*I do DNA plasmidial do clone pUFV1591. A canaleta 2, padrão de clivagem com as enzimas *Xba*I e *Nde*I do plasmídeo recombinante para obtenção do vetor de silenciamento gênico induzido por vírus, clone pUFV1713.

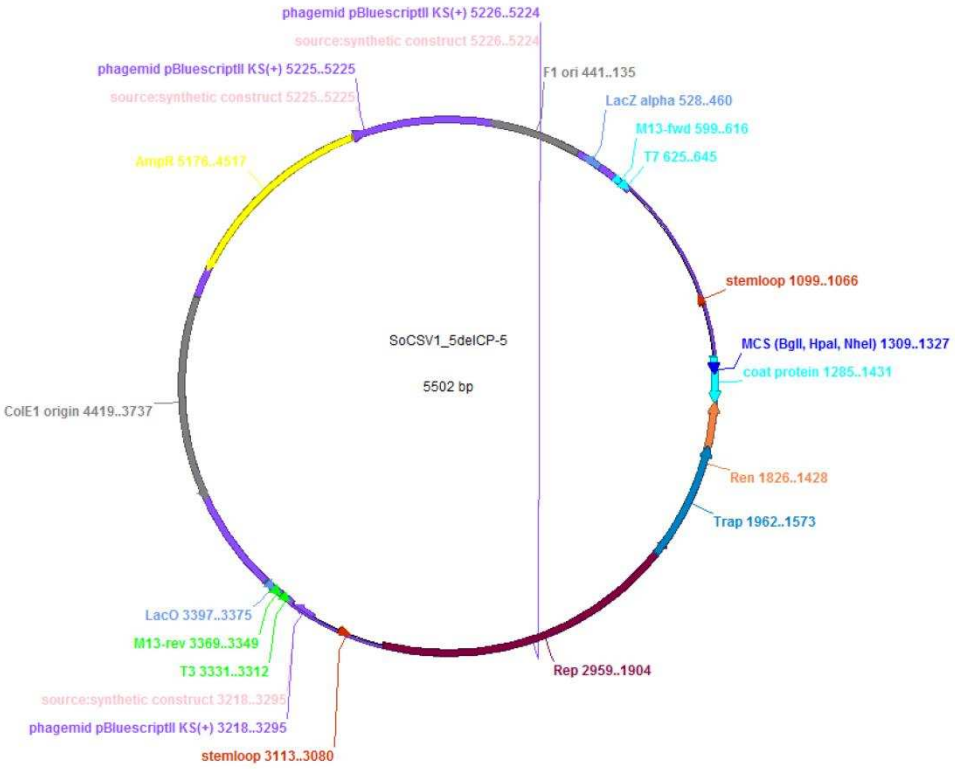


Figura 10: Mapa do vetor de silenciamento gênico induzido por vírus (VIGS: clone pUFV1713).

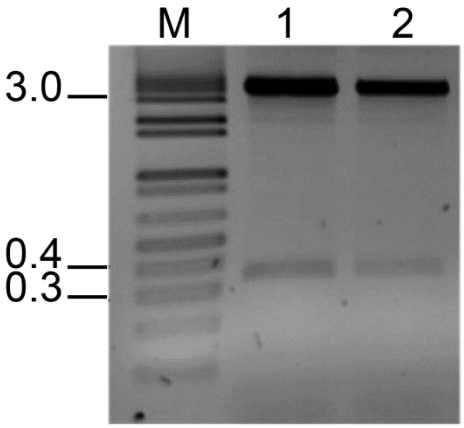


Figura 11: Diagnóstico molecular do clone pUFV1732 (VIGS-Chll). As canaletas 1-2 referem-se ao padrão eletroforético em gel de agarose (1%) proveniente da clivagem do DNA plasmidial do vetor de silenciamento viral (VIGS) contendo um fragmento do gene *ChlI* com as enzimas *Bgl*II e *Nhe*I (clone pUFV1732). A letra M refere-se ao marcador de tamanho de DNA em kpb.

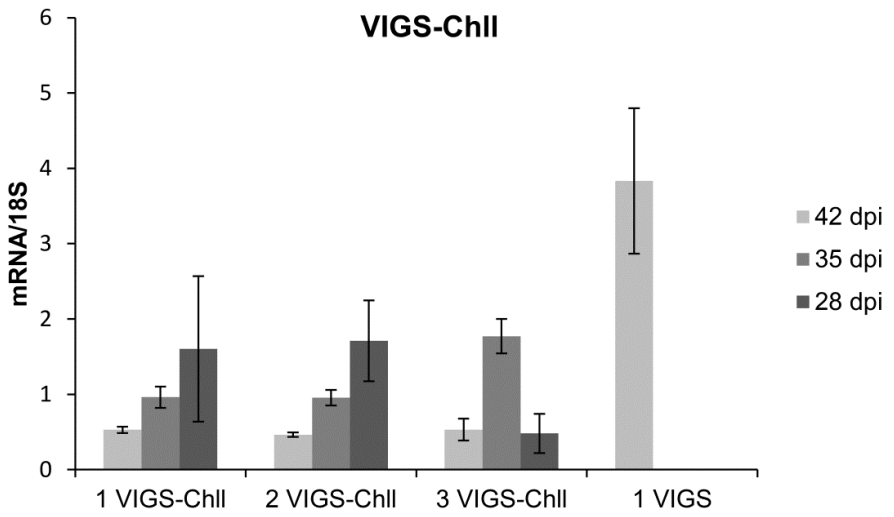


Figura 12: Quantificação por PCR em tempo real do mRNA do gene *magnesium chelatase subunit I* (*ChlI*) silenciado. Radículas de soja foram inoculadas com os vetores pUFV1732 (VIGS-Chll) + pUFV1592 (DNA-B) corresponde à planta 1 pUFV1732 (VIGS-Chll) + pUFV2573 (DNA-B) correspondente às plantas 2 e 3 e como controle foi utilizado a inoculação com o vetor pUFV1713 (VIGS) + pUFV1592 (DNA-B), correspondente ao vetor vazio para VIGS. Após 28 dias pós-inoculação (dpi), 35 dpi e 42dpi, o RNA total foi extraído de folhas infectadas e a expressão do gene *ChlI* foi determinada por RT-PCR quantitativo, utilizando o gene 18S como controle endógeno. As identificações "1 VIGS-Chll", "2 VIGS-Chll" e "3 VIGS-Chll" correspondem a quantificação a partir do cDNA sintetizado do RNA total respectivamente, da planta infectada identificada como planta 1, 2 e 3, conforme descrito acima. A identificação "1 VIGS" corresponde a quantificação a partir do cDNA sintetizado do RNA total de uma planta de soja bombardeada com o vetor de silenciamento viral vazio. Os resultados estão plotados com o intervalo de confiança com 95% de confiança.

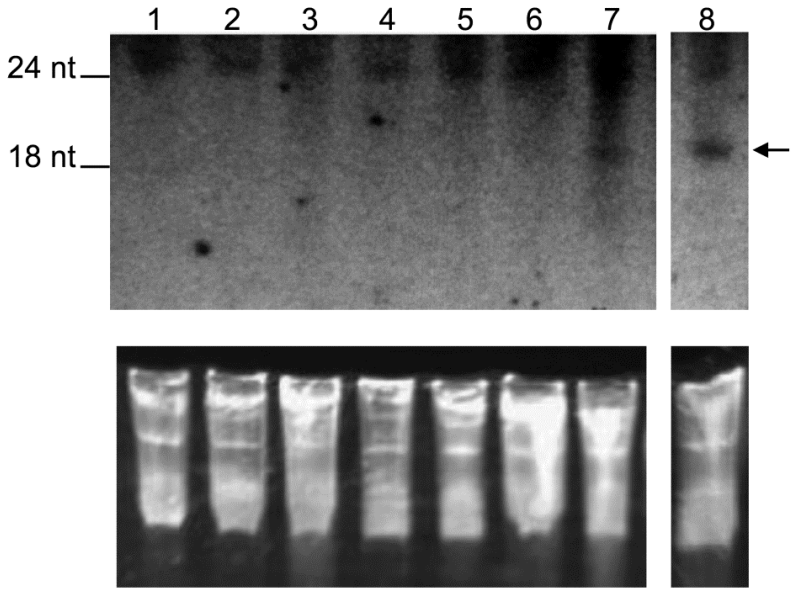


Figura 13: Detecção de siRNA em plantas silenciadas para o gene Chll. A indução do silenciamento do gene ChlI *(magnesium chelatase subunit I)* em plantas de soja inoculadas com pUFV1732 (VIGS-Chll) + pUFV1592 (DNA B) e pUFV1732 (VIGS-Chll) + pUFV2573 (DNA B) foi monitorada pela expressão de siRNA em planta de soja (canaletas 7 e 8, respectivamente), detectada por northern blots. A canaleta 1 corresponde ao RNA total de uma planta de soja inoculada com partículas de tungstênio sem DNA viral, controle negativo. A canaleta 2 corresponde ao RNA total de uma planta de soja inoculada com o vetor de silenciamento viral vazio pUFV1713 (VIGS vazio) + pUFV2573, 28 dias pós-inoculação (dpi). A canaleta 3 corresponde ao RNA total de uma planta de soja bombardeada com pUFV1713 (VIGS vazio) + pUFV2573, 35 dpi. A canaleta 4 corresponde ao RNA total de uma planta de soja inoculada com pUFV1713 (VIGS vazio) + pUFV2573, 42 dpi. A canaleta 5 corresponde ao RNA total de uma planta de soja inoculada com o vetor de silenciamento gênico pUFV1732 (VIGS-Chll) + pUFV1592 (DNA B), 28 dpi. A canaleta 6 corresponde ao RNA total de uma planta de soja bombardeada com o vetor de silenciamento gênico pUFV1732 (VIGS-Chll) + pUFV1592 (DNA B), 35 dpi. A canaleta 7 corresponde ao RNA total de uma planta de soja inoculada com o vetor de silenciamento gênico pUFV1732 (VIGS-Chll) + pUFV1592 (DNA B), 42 dpi. A canaleta 8 corresponde ao RNA total de uma planta de soja inoculada com o vetor de silenciamento gênico pUFV1732 (VIGS-Chll) + pUFV2573 (DNA B), 28 dpi. Observa-se o aparecimento de siRNA de 21 nt (indicado pela seta) nas canaletas 7 e 8. O gel de baixo corresponde ao gel de poliacrilamida 15% do RNA total corado com brometo de etídio.

RESUMO

VETOR VIRAL E USO DESTE PARA SILENCIAMENTO GÊNICO EM PLANTAS LEGUMINOSAS. A presente invenção refere-se a um vetor de silenciamento viral baseado na modificação do genoma do vírus *Soybean chlorotic spot virus* (SoCSV) para induzir o silenciamento sistêmico de genes endógenos de plantas pertencentes à família *Fabaceae*. Tal tecnologia oferece grandes vantagens por ser rápida e permitir a análise de genes em larga escala, possibilitando um avanço no conhecimento básico de funções gênicas desconhecidas e na consequente seleção de genes de interesse para o melhoramento genético de leguminosas no Brasil.

LISTAGEM DE SEQUÊNCIAS

<110> UNIVERSIDADE FEDERAL DE VIÇOSA
FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE MINAS GERAIS
<120> VETOR VIRAL E USO DESTE PARA SILENCIAMENTO GÊNICO EM PLANTAS LEGUMINOSAS
<130> 01 - 2015
<160> 1

<210> **SEQ ID NO:1**
<211> 2561
<212> DNA
<213> DNA-A de Soybean Chlorotic spot virus modificado

<400> 1
cgacttagct ccctgtatgt ttggatggaa atgtgctgac cttgttgggg aaaccaagtc 60
gaacaatctg ttatttgtgc agcagaactt cccttcgaat tgaagcagga catggagatg 120
aggctgccca tcgtcgtgaa gttctcttgc tatcttgata aacttcttac tgctaggtgt 180
ctgcaacgag ataatttgtt caagagcagc ctcttttgat aaggagcact ggggatatgt 240
gaggaaatag tttttggcat ttatacggaa acgctttggt ggtggcattc ttgtaaataa 300
aacgatgtca ccaattcgct ccgctctaaa actctatagc aattggtgaa ttggtgacaa 360
tatatactag aaccctcaat agaacttcta atcctgttcg cacacgtggc ggccatccgc 420
tataatatta ccggatggcc gcgcgatttt ttatggccct tttcttttgg gctcgttctt 480
ttggaccgag tgtatttgaa ttaaagtaaa gttattcccc tgtccaatga aaatttgtct 540
gagtggtcta gataagccca acttggtacc caagttggtt ctaacggtta tatattcaag 600
aataactgac taattcttta attcaaaatg gttaagcggg atgccccatg gagatctgtt 660
aacagctagc cggaaaacgc attattattg tacatggcat gtactcatgc ctctaacccc 720
gtgtatgcaa ctctaaaaat tcggatctat ttttatgatt cgataaccaa ttaataaaat 780
ttgaatttta ttacatgatt ctcaagtaca taatttacat atgatttgtc tgttgcgaaa 840
cgaacagctc ttattacatt gttaagcgag ataacaccta ctcgctctaa gtacagcatg 900
actaggaatt taaatctatt taaatatgtc gtcccagaag ctctcaggga tatcgtccag 960
acttggaaat tgagatatgc cttgtggaga gccaatgctc tcctcaggtt gtggttgaac 1020
ctgatctgga cgtggtatat tctggtcgtt gtgtgtagtg gatcctctac gtggttcatc 1080
ttgaaataga ggggatttga tatctcccag atatagacgc cattctctgc ctgacgtgca 1140
gtgatgagtt cccctgtgcg tgaatccatg tcctgcgcag ttaagatgga cgtatatgga 1200
gcacccgcaa tctatatcaa ttcgtcttct cctgatagcc ctcttttag ctgctctgtg 1260
ttgaactttg atagaggggg gagttgagga agacgaattt cgcattgtga agggtccaac 1320
ctttcaatcc tgcgttttcc tctttgtcga ggaaagattt atagctggag ccctcgccag 1380
gattgcaaag tataatgcaa ggaacccctc ctttaatttg aactggcttt ccgtatttgc 1440
agtttgattg ccagtctttt tgagcaccta tcaattcctt ccaatgcttt aactttagat 1500
attgcggtgc gatgtcatct attacgttgt actgcacatc atttgaatag acccgcgaat 1560
tgaaatcgag gtgaccggac aagtagttat gggaccctag cgcacgtgcc cacatggtct 1620
tccccgttcg agagtcaccc tcgattatta tactaatagg tctttccggc cgcgcagcgg 1680
cacctctccc aaaataatca ttgacccaat cttgcatctc ctgtggaact aaagtgaacg 1740
aggagaggtg aaacggagga gcccatggct ccggagcctt agcaaagata cgctctaggt 1800
tagagcgtat gttgtggtga tgaaggacga agtccttcgg ctgttcttct tttaatattt 1860
gaagggctct ttctaggttg tctgcgttca gaaccttggc gtatgactcg tttgcagtct 1920
gttgaccacc tcttgcagaa cgagcgtcga cctggaattc tccccattca atggtgtctc 1980
catccttgtc gacgtaggac ttgacgtcgg agctcgactt agctccctgt atgtttggat 2040
ggaaatgtgc tgaccttgtt ggggaaacca agtcgaacaa tctgttattt gtgcagcaga 2100
acttcccttc gaattgaagc aggacatgga gatgaggctg cccatcgtcg tgaagttctc 2160
ttgctatctt gataaacttc ttactgctag gtgtctgcaa cgagataatt tgttcaagag 2220
cagcctcttt tgataaggag cactggggat atgtgaggaa atagtttttg gcatttatac 2280
ggaaacgctt tggtggtggc attcttgtaa ataaaacgat gtcaccaatt cgctccgctc 2340
taaaactcta tagcaattgg tgaattggtg acaatatata ctagaaccct caatagaact 2400
tctaatcctg ttcgcacacg tggcggccat ccgctataat attaccggat ggccgcgcga 2460
ttttttatgg ccctttttctt ttgggctcgt tcttttggac cgagtgtatt tgaattaaag 2520
taaagttatt cccctgtcca atgaaaattt gtctgagtgg t 2561

<210> **SEQ ID N°2**
<211> 5503
<212> DNA
<213> DNA-A de Soybean Chlorotic spot virus modificado clonado em plasmídeo

<400> 2
cgacttagct ccctgtatgt ttggatggaa atgtgctgac cttgttgggg aaaccaagtc 60
gaacaatctg ttatttgtgc agcagaactt cccttcgaat tgaagcagga catggagatg 120
aggctgccca tcgtcgtgaa gttctcttgc tatcttgata aacttcttac tgctaggtgt 180
ctgcaacgag ataatttgtt caagagcagc ctcttttgat aaggagcact ggggatatgt 240
gaggaaatag tttttggcat ttatacggaa acgctttggt ggtggcattc ttgtaaataa 300
aacgatgtca ccaattcgct ccgctctaaa actctatagc aattggtgaa ttggtgacaa 360
tatatactag aaccctcaat agaacttcta atcctgttcg cacacgtggc ggccatccgc 420
tataatatta ccggatggcc gcgcgatttt ttatggccct tttcttttgg gctcgttctt 480
ttggaccgag tgtatttgaa ttaaagtaaa gttattcccc tgtccaatga aaatttgtct 540
gagtggtcta gataagccca acttggtacc caagttggtt ctaacggtta tatattcaag 600
aataactgac taattcttta attcaaaatg gttaagcggg atgccccatg gagatctgtt 660

```
aacagctagc cggaaaacgc attattattg tacatggcat gtactcatgc ctctaacccc 720
gtgtatgcaa ctctaaaaat tcggatctat ttttatgatt cgataaccaa ttaataaaat 780
ttgaatttta ttacatgatt ctcaagtaca taatttacat atgatttgtc tgttgcgaaa 840
cgaacagctc ttattacatt gttaagcgag ataacaccta ctcgctctaa gtacagcatg 900
actaggaatt taaatctatt taaatatgtc gtcccagaag ctctcaggga tatcgtccag 960
acttggaaat tgagatatgc cttgtggaga gccaatgctc tcctcaggtt gtggttgaac 1020
ctgatctgga cgtggtatat tctggtcgtt gtgtgtagtg gatcctctac gtggttcatc 1080
ttgaaataga ggggatttga tatctcccag atatagacgc cattctctgc ctgacgtgca 1140
gtgatgagtt cccctgtgcg tgaatccatg tcctgcgcag ttaagatgga cgtatatgga 1200
gcacccgcaa tctatatcaa ttcgtcttct cctgatagcc ctctttttag ctgctctgtg 1260
ttgaactttg atagaggggg gagttgagga agacgaattt cgcattgtga agggtccaac 1320
ctttcaatcc tgcgttttcc tctttgtcga ggaaagattt atagctggag ccctcgccag 1380
gattgcaaag tataatgcaa ggaacccctc ctttaatttg aactggcttt ccgtatttgc 1440
agtttgattg ccagtctttt tgagcaccta tcaattcctt ccaatgcttt aactttagat 1500
attgcggtgc gatgtcatct attacgttgt actgcacatc atttgaatag acccgcgaat 1560
tgaaatcgag gtgaccggac aagtagttat gggaccctag cgcacgtgcc cacatggtct 1620
tccccgttcg agagtcaccc tcgattatta tactaatagg tctttccggc cgcgcagcgg 1680
cacctctccc aaaataatca ttgacccaat cttgcatctc ctgtggaact aaagtgaacg 1740
aggagaggtg aaacggagga gcccatggct ccggagcctt agcaaagata cgctctaggt 1800
tagagcgtat gttgtggtga tgaaggacga agtccttcgg ctgttcttct tttaatattt 1860
gaagggctct ttctaggttg tctgcgttca gaaccttggc gtatgactcg tttgcagtct 1920
gttgaccacc tcttgcagaa cgagcgtcga cctggaattc tccccattca atggtgtctc 1980
catccttgtc gacgtaggac ttgacgtcgg agctcgactt agctccctgt atgtttggat 2040
ggaaatgtgc tgaccttgtt ggggaaacca agtcgaacaa tctgttattt gtgcagcaga 2100
acttcccttc gaattgaagc aggacatgga gatgaggctg cccatcgtcg tgaagttctc 2160
ttgctatctt gataaacttc ttactgctag gtgtctgcaa cgagataatt tgttcaagag 2220
cagcctcttt tgataaggag cactggggat atgtgaggaa atagtttttg gcatttatac 2280
ggaaacgctt tggtggtggc attcttgtaa ataaaacgat gtcaccaatt cgctccgctc 2340
taaaactcta tagcaattgg tgaattggtg acaatatata ctagaaccct caatagaact 2400
tctaatcctg ttcgcacacg tggcggccat ccgctataat attaccggat ggccgcgcga 2460
ttttttatgg ccctttttct ttgggctcgt tcttttggac cgagtgtatt tgaattaaag 2520
taaagttatt cccctgtcca atgaaaattt gtctgagtgg tctagaacta gtggatcccc 2580
cgggctgcag gaattcgata tcaagcttat cgataccgtc gacctcgagg gggggcccgg 2640
tacccagctt ttgttccctt tagtgagggt taattgcgcg cttggcgtaa tcatggtcat 2700
agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa 2760
gcataaagtg taaagcctgg ggtgcctaat gagtgagcta actcacatta attgcgttgc 2820
gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc 2880
aacgcgcggg gagaggcggt ttgcgtattg ggcgctcttc cgcttcctcg ctcactgact 2940
cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag gcggtaatac 3000
ggttatccac agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa 3060
aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgccccctg 3120
acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa 3180
gataccaggc gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc 3240
ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac 3300
gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac 3360
cccccgttca gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg 3420
taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt 3480
atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga 3540
cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct 3600
cttgatccgg caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga 3660
ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg 3720
ctcagtggaa cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct 3780
tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt 3840
aaacttggtc tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc 3900
tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg 3960
gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgctca ccggctccag 4020
atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggt cctgcaactt 4080
tatccgcctc catccagtct attaattgtt gccgggaagc tagagtaagt agttcgccag 4140
ttaatagttt gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt 4200
ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatccccca 4260
tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat cgttgtcaga agtaagttgg 4320
ccgcagtgtt atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat 4380
ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgta 4440
tgcggcgacc gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca 4500
gaactttaaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct 4560
taccgctgtt gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat 4620
cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa 4680
agggaataag ggcgacacgg aaatgttgaa tactcatact cttcctttttt caatattatt 4740
gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa 4800
ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctaaa ttgtaagcgt 4860
taatattttg ttaaaattcg cgttaaattt ttgttaaatc agctcatttt ttaaccaata 4920
ggccgaaatc ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt 4980
tgttccagtt tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg 5040
aaaaaccgtc tatcagggcg atggcccact acgtgaacca tcaccctaat caagtttttt 5100
ggggtcgagg tgccgtaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc 5160
ttgacgggga aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg 5220
cgctagggcg ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgccgcgct 5280
```

```
taatgcgccg ctacagggcg cgtcccattc gccattcagg ctgcgcaact gttgggaagg 5340
gcgatcggtg cgggcctctt cgctattacg ccagctggcg aaaggggat gtgctgcaag 5400
gcgattaagt tgggtaacgc cagggttttc ccagtcacga cgttgtaaaa cgacggccag 5460
tgagcgcgcg taatacgact cactataggg cgaattggag ct                    5502
```

<210> **SEQ ID N°3**
<211> 3351
<212> DNA
<213> DNA-B de Soybean Chlorotic spot virus modificado

<400> 3

```
atcaaccgaa tgctttgctg tggagagctt taacttccct ttgaatttcg cgaaatgtgt 60
tctttgatgg acatttgtgt cgcacactct atagtataat ttccagggaa ttgggtcttt 120
gagcgaaaaa aaagacgacg agaaatagtg gagatctatg ttacatctaa tgggaaacgt 180
ccatgacgct tgcaatgact cattgtccgt cattctcttg tcatgaatct ccactattac 240
tgatcctgtt gcgttgatag gaacctgttg tctgtactca atcacacaat ggtcgatttt 300
catacaacta cgactgattc tagcagttaa ctgagacgct gtggacggaa attgcaatac 360
tatctcagtt aggtcatgag aaagttgata ttcatcacgc tgagactcta cgtaattgaa 420
aactgcgggt ggataagcta actgagactc catagaacac aataataata aataaaaatg 480
gccgcgcagc ggaatggcaa cgacgaactg atcaaggaac aaggaaatgt ttttattgct 540
gtgaaaacag cagctgacta aaggaaataa gaagttgctt tcttattgaa tatgtgattc 600
ttgtttggtt taaatagaca tgaccatgat cgtacaacct aattattaaa tattatgtcc 660
tcacatattc cattcatctt ttatttacaa agaaaaacaa ataaaacaat gatattgtta 720
agaagtctgg cagtggcatg tttgtaaata aaacgatgtc accaattgag ctccgctcta 780
aaactctata gcaatcggtg aattggtgac aatatatact agaaccctct atagaacttc 840
taatcctgtt cgcacacgtg gcggccatcc gctataatat taccggatgg ccgcgcgatt 900
ttttaccctt ttgacgtggc gctctctctg ccgtccgttt attttaactt caattcagtt 960
tagcgcattt ttgagttccg ctcacacgta tcgtttttg cgcccgtgtg acatgtgcgc 1020
taatgtttaa ctttttttgac tagatttatt ggagaccgtt ggataaatct aactgatatt 1080
ttgaacaatc aacggtcaca ttttaacacc atgggccatt atttgttaag ctgacatctt 1140
taactccatt gtactatgta ttgtgacttc gaccaattaa attgtttttt gagagtcaat 1200
ttaattgctg tgttcattct gttgtgttat atatctgtgg acaatgttat ggaacaatac 1260
atatttcata aacatttgat catttaccta tgaatttcac cagaaataga cgtggtttga 1320
caactacttc gcgtagaagt agttataggt atcctttgtc tagacgatcg tatggtgtga 1380
aaagaaacaa tggtagacgt gggccgaata ttccaaacaa ggcccatgaa gatggaaaga 1440
tgtcaggtca atgcatacat gaaaatcagt ttggtccaga atttgttatg gcccataatt 1500
cacccatttc aactttcatt acatacccta ctctgggtaa aactgactcc agtcgatcaa 1560
ggtcgtatat taagttaaaa cgtctacgtt ttaagggaac tgtgaaagtt gaacgtctta 1620
ccccgaatgt taacatggac gggctacatc caaagataga aggagtcttt tcaatggtgg 1680
ttgtggtcga tcgaaaacca catttaagcc cgtccggatg tctccacaca ttcgatgaat 1740
tgtttggtgc caggatacac agtcatggaa atttagctat tactgcctcg ttgaaagacc 1800
gattctatat acgccatgtg tttaaacgcg tgatgtctgt cgagaaagat agcacgatgg 1860
tggatattga agggacgaca acattatcta ataggcgttt taattgttgg gctagtttta 1920
gagattttga acgtgattca tgtaacggtg tttatgccaa cattagcaaa aacgccttgt 1980
taatttacta ttgctggatg tcagataatg tgtctaaggc atctacattt gtatcatttg 2040
atctggaata tattggttga gtgataaatt gtgtttaaac ctgtcaagat ttttattaaa 2100
cgtcaaatga taatttgcac acgctattca aaaatgcaca aagacattat ttatttaaga 2160
cacttgggct cggaaggaat acagttgctt ttaatacatt cttgggccgt tgttctcacc 2220
agctcgttca attgggccat tgatattgtt atactggact ctgccctcct tgcacctact 2280
attgatgctg attcacccgg gtctaagaca ctgctgtcta atctatgcag ttgtctatat 2340
ggatgggttg tgttctcaac ctcagagtcc gcatcggttg aactgactcc tatggtgctt 2400
cttgtggccc atgtttctcc cggtttgatc gttattgggc tgtgtatgcc atatctcgtt 2460
gtggatgcgg acttgatcaa tttccgttca taatgaccat atcccacatg agaaaaatcg 2520
acatcgctct ccttgaactg tttcgataat attttaactg ttggagccct gaaagggata 2580
tcaaccgaat gctttgctgt ggagagcttt aacttccctt tgaatttcgc gaaatgtgtt 2640
ctttgatgga catttgtgtc gcacactcta tagtataatt tccagggaat tgggtctttg 2700
agcgaaaaaa aagacgacga gaaatagtgg agatctatgt tacatctaat gggaaacgtc 2760
catgacgctt gcaatgactc attgtccgtc attctcttgt catgaatctc cactattact 2820
gatcctgttg cgttgatagg aacctgttgt ctgtactcaa tcacacaatg gtcgatttc 2880
atacaactac gactgattct agcagttaac tgagacgctg tggacggaaa ttgcaatact 2940
atctcagtta ggtcatgaga aagttgatat tcatcacgct gagactctac gtaattgaaa 3000
actgcgggtg gataagctaa ctgagactcc atagaacaca ataataataa ataaaaatgg 3060
ccgcgcagcg gaatggcaac gacgaactga tcaaggaaca aggaaatgtt tttattgctg 3120
tgaaaacagc agctgactaa aggaaataag aagttgcttt cttattgaat atgtgattct 3180
tgtttggttt aaatagacat gaccatgatc gtacaaccta attattaaat attatgtcct 3240
cacatattcc attcatcttt tatttacaaa gaaaaacaaa taaaacaatg atattgttaa 3300
gaagtctggc agtggcatgt ttgtaaataa aacgatgtca ccaattgagc t           3351
```

<210> **SEQ ID N°4**
<211> 6250
<212> DNA
<213> DNA-B de Soybean Chlorotic spot virus modificado clonado em plasmídeo

<400> 4
atcaaccgaa tgctttgctg tggagagctt taacttccct ttgaatttcg cgaaatgtgt 60
tctttgatgg acatttgtgt cgcacactct atagtataat ttccagggaa ttgggtcttt 120
gagcgaaaaa aaagacgacg agaaatagtg gagatctatg ttacatctaa tgggaaacgt 180
ccatgacgct tgcaatgact cattgtccgt cattctcttg tcatgaatct ccactattac 240
tgatcctgtt gcgttgatag gaacctgttg tctgtactca atcacacaat ggtcgatttt 300
catacaacta cgactgattc tagcagttaa ctgagacgct gtggacggaa attgcaatac 360
tatctcagtt aggtcatgag aaagttgata ttcatcacgc tgagactcta cgtaattgaa 420
aactgcgggt ggataagcta actgagactc catagaacac aataataata aataaaaatg 480
gccgcgcagc ggaatggcaa cgacgaactg atcaaggaac aaggaaatgt ttttattgct 540
gtgaaaacag cagctgacta aaggaaataa gaagttgctt tcttattgaa tatgtgattc 600
ttgtttggtt taaatagaca tgaccatgat cgtacaacct aattattaaa tattatgtcc 660
tcacatattc cattcatctt ttatttacaa agaaaaacaa ataaaacaat gatattgtta 720
agaagtctgg cagtggcatg tttgtaaata aaacgatgtc accaattgag ctccgctcta 780
aaactctata gcaatcggtg aattggtgac aatatatact agaaccctct atagaacttc 840
taatcctgtt cgcacacgtg gcggccatcc gctataatat taccggatgg ccgcgcgatt 900
ttttaccctt ttgacgtggc gctctctctg ccgtccgttt attttaactt caattcagtt 960
tagcgcattt ttgagttccg ctcacacgta tcgtttttttg cgcccgtgtg acatgtgcgc 1020
taatgtttaa ctttttttgac tagatttatt ggagaccgtt ggataaatct aactgatatt 1080
ttgaacaatc aacggtcaca ttttaacacc atgggccatt atttgttaag ctgacatctt 1140
taactccatt gtactatgta ttgtgacttc gaccaattaa attgtttttt gagagtcaat 1200
ttaattgctg tgttcattct gttgtgttat atatctgtgg acaatgttat ggaacaatac 1260
atatttcata aacatttgat catttaccta tgaatttcac cagaaataga cgtggtttga 1320
caactacttc gcgtagaagt agttataggt atcctttgtc tagacgatcg tatggtgtga 1380
aaagaaacaa tggtagacgt gggccgaata ttccaaacaa ggcccatgaa gatggaaaga 1440
tgtcaggtca atgcatacat gaaaatcagt ttggtccaga atttgttatg gcccataatt 1500
cacccatttc aactttcatt acatacccta ctctgggtaa aactgactcc agtcgatcaa 1560
ggtcgtatat taagttaaaa cgtctacgtt ttaagggaac tgtgaaagtt gaacgtctta 1620
ccccgaatgt taacatggac gggctacatc caaagataga aggagtcttt tcaatggtgg 1680
ttgtggtcga tcgaaaacca catttaagcc cgtccggatg tctccacaca ttcgatgaat 1740
tgtttggtgc caggatacac agtcatggaa atttagctat tactgcctcg ttgaaagacc 1800
gattctatat acgccatgtg tttaaacgcg tgatgtctgt cgagaaagat agcacgatgg 1860
tggatattga agggacgaca acattatcta ataggcgttt taattgttgg gctagtttta 1920
gagattttga acgtgattca tgtaacggtg tttatgccaa cattagcaaa aacgccttgt 1980
taatttacta ttgctggatg tcagataatg tgtctaaggc atctacattt gtatcatttg 2040
atctggaata tattggttga gtgataaatt gtgtttaaac ctgtcaagat ttttattaaa 2100
cgtcaaatga taatttgcac acgctattca aaaatgcaca aagacattat ttatttaaga 2160
cacttgggct cggaaggaat acagttgctt ttaatacatt cttgggccgt tgttctcacc 2220
agctcgttca attgggccat tgatattgtt atactggact ctgccctcct tgcacctact 2280
attgatgctg attcacccgg gtctaagaca ctgctgtcta atctatgcag ttgtctatat 2340
ggatgggttg tgttctcaac ctcagagtcc gcatcggttg aactgactcc tatggtgctt 2400
cttgtggccc atgtttctcc cggtttgatc gttattgggc tgtgtatgcc atatctcgtt 2460
gtggatgcgg acttgatcaa tttccgttca taatgaccat atcccacatg agaaaaatcg 2520
acatcgctct ccttgaactg tttcgataat attttaactg ttggagccct gaaagggata 2580
tcaaccgaat gctttgctgt ggagagcttt aacttccctt tgaatttcgc gaaatgtgtt 2640
ctttgatgga catttgtgtc gcacactcta tagtataatt tccagggaat tgggtctttg 2700
agcgaaaaaa aagacgacga gaaatagtgg agatctatgt tacatctaat gggaaacgtc 2760
catgacgctt gcaatgactc attgtccgtc attctcttgt catgaatctc cactattact 2820
gatcctgttg cgttgatagg aacctgttgt ctgtactcaa tcacacaatg gtcgattttc 2880
atacaactac gactgattct agcagttaac tgagacgctg tggacggaaa ttgcaatact 2940
atctcagtta ggtcatgaga aagttgatat tcatcacgct gagactctac gtaattgaaa 3000
actgcgggtg gataagctaa ctgagactcc atagaacaca ataataataa ataaaaatgg 3060
ccgcgcagcg gaatggcaac gacgaactga tcaaggaaca aggaaatgtt tttattgctg 3120
tgaaaacagc agctgactaa aggaaataag aagttgcttt cttattgaat atgtgattct 3180
tgtttggttt aaatagacat gaccatgatc gtacaaccta attattaaat attatgtcct 3240
cacatattcc attcatcttt tatttacaaa gaaaaacaaa taaaacaatg atattgttaa 3300
gaagtctggc agtggcatgt ttgtaaataa aacgatgtca ccaattgagc tccaattcgc 3360
cctatagtga gtcgtattac gcgcgctcac tggccgtcgt tttacaacgt cgtgactggg 3420
aaaaccctgg cgttacccaa cttaatcgcc ttgcagcaca tccccctttc gccagctggc 3480
gtaatagcga agaggcccgc accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg 3540
aatgggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg 3600
tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc 3660
tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggggctccct ttagggttcc 3720
gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat ggttcacgta 3780
gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta 3840
atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc tattcttttg 3900
atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg atttaacaaa 3960
aatttaacgc gaattttaac aaaatattaa cgcttacaat ttacgtggca cttttcgggg 4020
aaatgtgcgc ggaacccta tttgtttatt tttctaaata cattcaaata tgtatccgct 4080
catgagacaa taaccctgat aaatgcttca ataatattga aaaaggaaga gtatgagtat 4140
tcaacatttc cgtgtcgccc ttattccctt ttttgcggca ttttgccttc ctgtttttgc 4200
tcacccagaa acgctggtga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg 4260

```
ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg 4320
ttttccaatg atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga 4380
cgccgggcaa gagcaactcg gtcgccgcat acactattct cagaatgact tggttgagta 4440
ctcaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc 4500
tgccataacc atgagtgata acactgcggc caacttactt ctgacaacga tcggaggacc 4560
gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg 4620
ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc 4680
aatggcaaca acgttgcgca aactattaac tggcgaacta cttactctag cttcccggca 4740
acaattaata gactggatgg aggcggataa agttgcagga ccacttctgc gctcggccct 4800
tccggctggc tggtttattg ctgataaatc tggagccggt gagcgtgggt ctcgcggtat 4860
cattgcagca ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg 4920
gagtcaggca actatggatg aacgaaatag acagatcgct gagataggtg cctcactgat 4980
taagcattgg taactgtcag accaagttta ctcatatata ctttagattg atttaaaact 5040
tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaaat 5100
cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc 5160
ttcttgagat ccttttttc tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct 5220
accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga aggtaactgg 5280
cttcagcaga gcgcagatac caaatactgt tcttctagtg tagccgtagt taggccacca 5340
cttcaagaac tctgtagcac cgcctacata cctcgctctg ctaatcctgt taccagtggc 5400
tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga 5460
taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct tggagcgaac 5520
gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca cgcttcccga 5580
agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag 5640
ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg 5700
acttgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga aaaacgccag 5760
caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgctcaca tgttctttcc 5820
tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc 5880
tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aagagcgccc 5940
aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag 6000
gtttcccgac tggaaagcgg gcagtgagcg caacgcaatt aatgtgagtt agctcactca 6060
ttaggcaccc caggctttac actttatgct cccggctcgt atgttgtgtg gaattgtgag 6120
cggataacaa tttcacacag gaaacagcta tgaccatgat tacgccaagc gcgcaattaa 6180
ccctcactaa agggaacaaa agctgggtac cgggcccccc ctcgaggtcg acggtatcga 6240
taagcttgat                                                       6250
```

<210> **SEQ ID N°5**
<211> 3167
<212> DNA
<213> DNA-B de Soybean Chlorotic spot virus modificado

<400> 5

```
ccgctctaaa actctatagc aatcggtgaa ttggtgacaa tatatactag aacccctctat 60
agaacttcta atcctgttcg cacacgtggc ggccatccgc tataatatta ccggatggcc 120
gcgcgatttt ttacccttttt gacgtggcgc tctctctgcc gtccgtttat tttaacttca 180
attcagttta gcgcatttttt gagttccgct cacacgtatc gttttttgcg cccgtgtgac 240
atgtgcgcta atgtttaact tttttgacta gatttattgg agaccgttgg ataaatctaa 300
ctgatatttt gaacaatcaa cggtcacatt ttaacaccat gggccattat ttgttaagct 360
gacatcttta actccattgt actatgtatt gtgacttcga ccaattaaat tgtttttga 420
gagtcaattt aattgctgtg ttcattctgt tgtgttatat atctgtggac aatgttatgg 480
aacaatacat atttcataaa catttgatca tttacctatg aatttcacca gaaatagacg 540
tggtttgaca actacttcgc gtagaagtag ttataggtat cctttgtcta gacgatcgta 600
tggtgtgaaa agaaacaatg gtagacgtgg gccgaatatt ccaaacaagg cccatgaaga 660
tggaaagatg tcaggtcaat gcatacatga aaatcagttt ggtccagaat ttgttatggc 720
ccataattca cccatttcaa ctttcattac ataccctact ctgggtaaaa ctgactccag 780
tcgatcaagg tcgtatatta agttaaaacg tctacgtttt aagggaactg tgaaagttga 840
acgtcttacc ccgaatgtta acatggacgg gctacatcca aagatagaag gagtcttttc 900
aatggtggtt gtggtcgatc gaaaaccaca tttaagcccg tccggatgtc tccacacatt 960
cgatgaattg tttggtgcca ggatacacag tcatggaaat ttagctatta ctgcctcgtt 1020
gaaagaccga ttctatatac gccatgtgtt taaacgcgtg atgtctgtcg agaaagatag 1080
cacgatggtg gatattgaag ggacgacaac attatctaat aggcgtttta attgttgggc 1140
tagttttaga gattttgaac gtgattcatg taacggtgtt tatgccaaca ttagcaaaaa 1200
cgccttgtta atttactatt gctggatgtc agataatgtg tctaaggcat ctacatttgt 1260
atcatttgat ctggaatata ttggttgagt gataaattgt gtttaaacct gtcaagattt 1320
ttattaaacg tcaaatgata atttgcacac gctattcaaa aatgcacaaa gacattattt 1380
atttaagaca cttgggctcg gaaggaatac agttgctttt aatacattct tgggccgttg 1440
ttctcaccag ctcgttcaat tgggccattg atattgttat actggactct gccctccttg 1500
cacctactat tgatgctgat tcacccgggt ctaagacact gctgtctaat ctatgcagtt 1560
gtctatatgg atgggttgtg ttctcaacct cagagtccgc atcggttgaa ctgactccta 1620
tggtgcttct tgtggcccat gtttctcccg gtttgatcgt tattgggctg tgtatgccat 1680
atctcgttgt ggatgcggac ttgatcaatt tccgttcata atgaccatat cccacatgag 1740
aaaaatcgac atcgctctcc ttgaactgtt tcgataatat tttaactgtt ggagccctga 1800
aagggatatc aaccgaatgc tttgctgtgg agagctttaa cttccctttg aatttcgcga 1860
aatgtgttct ttgatggaca tttgtgtcgc acactctata gtataatttc cagggaattg 1920
ggtctttgag cgaaaaaaaa gacgacgaga aatagtggag atctatgtta catctaatgg 1980
gaaacgtcca tgacgcttgc aatgactcat tgtccgtcat tctcttgtca tgaatctcca 2040
ctattactga tcctgttgcg ttgataggaa cctgttgtct gtactcaatc acacaatggt 2100
```

```
cgattttcat acaactacga ctgattctag cagttaactg agacgctgtg gacggaaatt 2160
gcaatactat ctcagttagg tcatgagaaa gttgatattc atcacgctga gactctacgt 2220
aattgaaaac tgcgggtgga taagctaact gagactccat agaacacaat aataataaat 2280
aaaaatggcc gcgcagcgga atggcaacga cgaactgatc aaggaacaag gaaatgtttt 2340
tattgctgtg aaaacagcag ctgactaaag gaaataagaa gttgctttct tattgaatat 2400
gtgattcttg tttggtttaa atagacatga ccatgatcgt acaacctaat tattaaatat 2460
tatgtcctca catattccat tcatctttta tttacaaaga aaaacaaata aaacaatgat 2520
attgttaaga agtctggcag tggcatgttt gtaaataaaa cgatgtcacc aattgagctc 2580
ccgctctaaa actctatagc aatcggtgaa ttggtgacaa tatatactag aaccctctat 2640
agaacttcta atcctgttcg cacacgtggc ggccatccgc tataatatta ccggatggcc 2700
gcgcgatttt ttacccttttt gacgtggcgc tctctctgcc gtccgtttat tttaacttca 2760
attcagttta gcgcattttt gagttccgct cacacgtatc gttttttgcg cccgtgtgac 2820
atgtgcgcta atgtttaact tttttgacta gatttattgg agaccgttgg ataaatctaa 2880
ctgatatttt gaacaatcaa cggtcacatt ttaacaccat gggccattat ttgttaagct 2940
gacatcttta actccattgt actatgtatt gtgacttcga ccaattaaat tgtttttga 3000
gagtcaattt aattgctgtg ttcattctgt tgtgttatat atctgtggac aatgttatgg 3060
aacaatacat atttcataaa catttgatca tttacctatg aatttcacca gaaatagacg 3120
tggtttgaca actacttcgc gtagaagtag ttataggtat cctttgt               3167
```

<210> **SEQ ID N°6**
<211> 6108
<212> DNA
<213> DNA-B de Soybean Chlorotic spot virus modificado clonado em plasmídeo

<400> 6

```
ccgctctaaa actctatagc aatcggtgaa ttggtgacaa tatatactag aaccctctat 60
agaacttcta atcctgttcg cacacgtggc ggccatccgc tataatatta ccggatggcc 120
gcgcgatttt ttacccttttt gacgtggcgc tctctctgcc gtccgtttat tttaacttca 180
attcagttta gcgcattttt gagttccgct cacacgtatc gttttttgcg cccgtgtgac 240
atgtgcgcta atgtttaact tttttgacta gatttattgg agaccgttgg ataaatctaa 300
ctgatatttt gaacaatcaa cggtcacatt ttaacaccat gggccattat ttgttaagct 360
gacatcttta actccattgt actatgtatt gtgacttcga ccaattaaat tgtttttga 420
gagtcaattt aattgctgtg ttcattctgt tgtgttatat atctgtggac aatgttatgg 480
aacaatacat atttcataaa catttgatca tttacctatg aatttcacca gaaatagacg 540
tggtttgaca actacttcgc gtagaagtag ttataggtat cctttgtcta gacgatcgta 600
tggtgtgaaa agaaacaatg gtagacgtgg gccgaatatt ccaaacaagg cccatgaaga 660
tggaaagatg tcaggtcaat gcatacatga aaatcagttt ggtccagaat ttgttatggc 720
ccataattca cccatttcaa ctttcattac ataccctact ctgggtaaaa ctgactccag 780
tcgatcaagg tcgtatatta agttaaaacg tctacgtttt aagggaactg tgaaagttga 840
acgtcttacc ccgaatgtta acatggacgg gctacatcca aagatagaag gagtcttttc 900
aatggtggtt gtggtcgatc gaaaaccaca tttaagcccg tccggatgtc tccacacatt 960
cgatgaattg tttggtgcca ggatacacag tcatggaaat ttagctatta ctgcctcgtt 1020
gaaagaccga ttctatatac gccatgtgtt taaacgcgtg atgtctgtcg agaaagatag 1080
cacgatggtg gatattgaag ggacgacaac attatctaat aggcgtttta attgttgggc 1140
tagttttaga gattttgaac gtgattcatg taacggtgtt tatgccaaca ttagcaaaaa 1200
cgccttgtta atttactatt gctggatgtc agataatgtg tctaaggcat ctacatttgt 1260
atcatttgat ctggaatata ttggttgagt gataaattgt gtttaaacct gtcaagattt 1320
ttattaaacg tcaaatgata atttgcacac gctattcaaa aatgcacaaa gacattattt 1380
atttaagaca cttgggctcg gaaggaatac agttgctttt aatacattct tgggccgttg 1440
ttctcaccag ctcgttcaat tgggccattg atattgttat actggactct gccctccttg 1500
cacctactat tgatgctgat tcacccgggt ctaagacact gctgtctaat ctatgcagtt 1560
gtctatatgg atgggttgtg ttctcaacct cagagtccgc atcggttgaa ctgactccta 1620
tggtgcttct tgtggcccat gtttctcccg gtttgatcgt tattgggctg tgtatgccat 1680
atctcgttgt ggatgcggac ttgatcaatt tccgttcata atgaccatat cccacatgag 1740
aaaaatcgac atcgctctcc ttgaactgtt tcgataatat tttaactgtt ggagccctga 1800
aagggatatc aaccgaatgc tttgctgtgg agagctttaa cttccctttg aatttcgcga 1860
aatgtgttct ttgatggaca tttgtgtcgc acactctata gtataatttc cagggaattg 1920
ggtctttgag cgaaaaaaaa gacgacgaga aatagtggag atctatgtta catctaatgg 1980
gaaacgtcca tgacgcttgc aatgactcat tgtccgtcat tctcttgtca tgaatctcca 2040
ctattactga tcctgttgcg ttgataggaa cctgttgtct gtactcaatc acacaatggt 2100
cgattttcat acaactacga ctgattctag cagttaactg agacgctgtg gacggaaatt 2160
gcaatactat ctcagttagg tcatgagaaa gttgatattc atcacgctga gactctacgt 2220
aattgaaaac tgcgggtgga taagctaact gagactccat agaacacaat aataataaat 2280
aaaaatggcc gcgcagcgga atggcaacga cgaactgatc aaggaacaag gaaatgtttt 2340
tattgctgtg aaaacagcag ctgactaaag gaaataagaa gttgctttct tattgaatat 2400
gtgattcttg tttggtttaa atagacatga ccatgatcgt acaacctaat tattaaatat 2460
tatgtcctca catattccat tcatctttta tttacaaaga aaaacaaata aaacaatgat 2520
attgttaaga agtctggcag tggcatgttt gtaaataaaa cgatgtcacc aattgagctc 2580
ccgctctaaa actctatagc aatcggtgaa ttggtgacaa tatatactag aaccctctat 2640
agaacttcta atcctgttcg cacacgtggc ggccatccgc tataatatta ccggatggcc 2700
gcgcgatttt ttacccttttt gacgtggcgc tctctctgcc gtccgtttat tttaacttca 2760
attcagttta gcgcattttt gagttccgct cacacgtatc gttttttgcg cccgtgtgac 2820
atgtgcgcta atgtttaact tttttgacta gatttattgg agaccgttgg ataaatctaa 2880
ctgatatttt gaacaatcaa cggtcacatt ttaacaccat gggccattat ttgttaagct 2940
gacatcttta actccattgt actatgtatt gtgacttcga ccaattaaat tgtttttga 3000
gagtcaattt aattgctgtg ttcattctgt tgtgttatat atctgtggac aatgttatgg 3060
```

```
aacaatacat atttcataaa catttgatca tttacctatg aatttcacca gaaatagacg 3120
tggtttgaca actacttcgc gtagaagtag ttataggtat cctttgtcta gaactagtgg 3180
atccccggg ctgcaggaat tcgatatcaa gcttatcgat accgtcgacc tcgaggggg 3240
gcccggtacc cagcttttgt tccctttagt gagggttaat tgcgcgcttg gcgtaatcat 3300
ggtcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac aacatacgag 3360
ccgggagcat aaagtgtaaa gcctggggtg cctaatgagt gagctaactc acattaattg 3420
cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa 3480
tcggccaacg cgcggggaga ggcggtttgc gtattgggcg ctcttccgct tcctcgctca 3540
ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg 3600
taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc 3660
agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc 3720
cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac 3780
tataaagata ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc 3840
tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata 3900
gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc 3960
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca 4020
acccggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 4080
cgaggtatgt aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta 4140
gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg 4200
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggtttttt gtttgcaagc 4260
agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt 4320
ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa 4380
ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat 4440
atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga 4500
tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata actacgatac 4560
gggagggctt accatctggc cccagtgctg caatgatacc gcgagaccca cgctcaccgg 4620
ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggtcctg 4680
caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt 4740
cgccagttaa tagtttgcgc aacgttgttg ccattgctac aggcatcgtg gtgtcacgct 4800
cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga gttacatgat 4860
cccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta 4920
agttggccgc agtgttatca ctcatggtta tggcagcact gcataattct cttactgtca 4980
tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat 5040
agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac 5100
atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa 5160
ggatcttacc gctgttgaga tccagttcga tgtaacccac tcgtgcaccc aactgatctt 5220
cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg 5280
caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat 5340
attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt 5400
agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcca cctaaattgt 5460
aagcgttaat attttgttaa aattcgcgtt aaatttttgt taaatcagct catttttaa 5520
ccaataggcc gaaatcggca aaatccctta taaatcaaaa gaatagaccg agatagggtt 5580
gagtgttgtt ccagtttgga acaagagtcc actattaaag aacgtggact ccaacgtcaa 5640
agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac cctaatcaag 5700
ttttttgggg tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga gcccccgatt 5760
tagagcttga cggggaaagc cggcgaacgt ggcgagaaag gaagggaaga aagcgaaagg 5820
agcgggcgct agggcgctgg caagtgtagc ggtcacgctg cgcgtaacca ccacacccgc 5880
cgcgcttaat gcgccgctac agggcgcgtc ccattcgcca ttcaggctgc gcaactgttg 5940
ggaagggcga tcggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc 6000
tgcaaggcga ttaagttggg taacgccagg gttttcccag tcacgacgtt gtaaaacgac 6060
ggccagtgag cgcgcgtaat acgactcact atagggcgaa ttggagct              6108
```

<210> **SEQ ID NO:7**
<211> 23
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo SoCSVNcoI-R

<400> 7
ggagccatgg gctcctccgt ttc 23

<210> **SEQ ID NO:8**
<211>
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo SoCSVMCSF

<400> 8
gaaaccatgg agatctgtta acagctagcc ggaaaacgca ttattattg 49

<210> **SEQ ID NO:9**
<211> 33
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo GmaxChlIF

<400> 9
cgagaattca tgggtgttga agtcagagag cgt 33

<210> **SEQ ID NO:10**
<211> 33
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo GmaxChlIR

<400> 10
cgagaattca ccgtgcagga tgcgagatag aaa 33

<210> **SEQ ID NO:11**
<211> 21
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo ChlIF

<400> 11
gccttcgcgc tacttctctt c 21

<210> **SEQ ID NO:12**
<211> 20
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo ChlIR

<400> 12
aaacggctac cacatccaag 20

<210> **SEQ ID NO:13**
<211> 21
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo 18SR específico para o gene de soja 18S

<400> 13
ccttcaatgg atccatcgtt a

<210> **SEQ ID NO:14**
<211> 20
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo 18SF específico para o gene de soja 18S

<400> 14
aaacggctac cacatccaag 20

<210> **SEQ ID NO:15**
<211> 28
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo Chll149FWD

<400> 15
acgattgaca ttgagaaagc cctgactg 28

<210> **SEQ ID NO:16**
<211> 29
<212> DNA
<213> Sequência artificial

<220> primer_bind
<223> oligonucleotídeo Chll149RVS

<400> 16
gatgcagcag aatccaacaa cacatccac 29

<210> **SEQ ID NO:17**
<211> 2623
<212> DNA
<213> DNA-A de Soybean Chlorotic spot virus

<400> 17

ACCGGATGGC CGCGCGATTT TTTATGGCCC TTTTCTTTTG GGCTCGTTCT TTTGGACCGA 60

GTGTATTTGA ATTAAAGTAA AGTTATTCCC CTGTCCAATG AAAATTTGTC TGAGTGGTCT 120

AGATAAGCCC AACTTGGTAC CCAAGTTGGT TCTAACGGTT ATATATTCAA GAATAACTGA 180

CTAATTCTTT AATTCAAAAT GGTTAAGCGG GATGCCCCAT GGCGCCACAT GGCAGGGACC 240

TCAAAGGTAT CCCGCTCTAG CAATTTTTCA CCTCGTGGAG GAGGAGGTCC AAAAAACAAC 300

AGGGCCTCTG AATGGGTCAA CAGGCCCATG TACCGGAAGC CCAGGATATA TCGGATGTAT 360

AGAACCCCCG ATGTACCTAG AGGTTGTGAA GGCCCTTGTA AGGTCCAGTC CTTTGAACAG 420

CGACACGATA TATCCCATGT TGGTAAGGTG ATATGTGTGT CTGACGTGAC AAGAGGTAAT 480

GGTATTACCC ATCGTGTAGG GAAACGTTTT TGTGTTAAGT CCGTTTATAT TCTAGGTAAG 540

ATATGGATGG ACGAGAATAT CAAGTTGAAG AATCACACGA ACAGTGTGAT GTTTTGGTTG 600

GTTAGAGACA GGAGACCCTA TGGCACCCCT ATGGACTTTG GCCAAGTGTT CAACATGTTC 660

GACAATGAAC CCAGTACTGC TACGGTTAAG AACGATCTTC GTGATCGTTT TCAAGTCATG 720

CACCGGTTCT ATGCCAAAGT CACTGGTGGA CAATATGCGA GCAACGAACA GGCATTGGTC 780

AGGCGTTTTT GGAAGGTCAA CAATCATGTT GTCTACAACC ACCAAGAAGC AGGGAAATAC 840

GAGAATCATA CGGAAAACGC ATTATTATTG TACATGGCAT GTACTCATGC CTCTAACCCC 900

GTGTATGCAA CTCTAAAAAT TCGGATCTAT TTTTATGATT CGATAACCAA TTAATAAAAT 960

TTGAATTTTA TTACATGATT CTCAAGTACA TAATTTACAT ATGATTTGTC TGTTGCGAAA 1020

CGAACAGCTC TTATTACATT GTTAAGCGAG ATAACACCTA CTCGCTCTAA GTACAGCATG 1080

ACTAGGAATT TAAATCTATT TAAATATGTC GTCCCAGAAG CTCTCAGGGA TATCGTCCAG 1140

ACTTGGAAAT TGAGATATGC CTTGTGGAGA GCCAATGCTC TCCTCAGGTT GTGGTTGAAC 1200

CTGATCTGGA CGTGGTATAT TCTGGTCGTT GTGTGTAGTG GATCCTCTAC GTGGTTCATC 1260

TTGAAATAGA GGGGATTTGA TATCTCCCAG ATATAGACGC CATTCTCTGC CTGACGTGCA 1320

GTGATGAGTT CCCCTGTGCG TGAATCCATG TCCTGCGCAG TTAAGATGGA CGTATATGGA 1380

GCACCCGCAA TCTATATCAA TTCGTCTTCT CCTGATAGCC CTCTTTTTAG CTGCTCTGTG 1440

TTGAACTTTG ATAGAGGGGG GAGTTGAGGA AGACGAATTT CGCATTGTGA AGGGTCCAAC 1500

CTTTCAATCC TGCGTTTTCC TCTTTGTCGA GGAAAGATTT ATAGCTGGAG CCCTCGCCAG 1560

GATTGCAAAG TATAATGCAA GGAACCCCTC CTTTAATTTG AACTGGCTTT CCGTATTTGC 1620

AGTTTGATTG CCAGTCTTTT TGAGCACCTA TCAATTCCTT CCAATGCTTT AACTTTAGAT 1680

ATTGCGGTGC GATGTCATCT ATTACGTTGT ACTGCACATC ATTTGAATAG ACCCGCGAAT 1740

TGAAATCGAG GTGACCGGAC AAGTAGTTAT GGGACCCTAG CGCACGTGCC CACATGGTCT 1800

TCCCCGTTCG AGAGTCACCC TCGATTATTA TACTAATAGG TCTTTCCGGC CGCGCAGCGG 1860

CACCTCTCCC AAAATAATCA TTGACCCAAT CTTGCATCTC CTGTGGAACT AAAGTGAACG 1920

AGGAGAGGTG AAACGGAGGA GCCCATGGCT CCGGAGCCTT AGCAAAGATA CGCTCTAGGT 1980

TAGAGCGTAT GTTGTGGTGA TGAAGGACGA AGTCCTTCGG CTGTTCTTCT TTTAATATTT 2040

GAAGGGCTCT TTCTAGGTTG TCTGCGTTCA GAACCTTGGC GTATGACTCG TTTGCAGTCT 2100

GTTGACCACC TCTTGCAGAA CGAGCGTCGA CCTGGAATTC TCCCCATTCA ATGGTGTCTC 2160

CATCCTTGTC GACGTAGGAC TTGACGTCGG AGCTCGACTT AGCTCCCTGT ATGTTTGGAT 2220

GGAAATGTGC TGACCTTGTT GGGGAAACCA AGTCGAACAA TCTGTTATTT GTGCAGCAGA 2280

ACTTCCCTTC GAATTGAAGC AGGACATGGA GATGAGGCTG CCCATCGTCG TGAAGTTCTC 2340

TTGCTATCTT GATAAACTTC TTACTGCTAG GTGTCTGCAA CGAGATAATT TGTTCAAGAG 2400

CAGCCTCTTT TGATAAGGAG CACTGGGGAT ATGTGAGGAA ATAGTTTTTG GCATTTATAC 2460

GGAAACGCTT TGGTGGTGGC ATTCTTGTAA ATAAAACGAT GTCACCAATT CGCTCCGCTC 2520

TAAAACTCTA TAGCAATTGG TGAATTGGTG ACAATATATA CTAGAACCCT CAATAGAACT 2580

TCTAATCCTG TTCGCACACG TGGCGGCCAT CCGCTATAAT ATT 2623

<210> **SEQ ID NO:18**
<211> 2585
<212> DNA
<213> DNA-B de Soybean Chlorotic spot virus

<400> 18
ACCGGATGGC CGCGCGATTT TTTACCCTTT TGACGTGGCG CTCTCTCTGC CGTCCGTTTA 60
TTTTAACTTC AATTCAGTTT AGCGCATTTT TGAGTTCCGC TCACACGTAT CGTTTTTTGC 120
GCCCGTGTGA CATGTGCGCT AATGTTTAAC TTTTTTGACT AGATTTATTG GAGACCGTTG 180
GATAAATCTA ACTGATATTT TGAACAATCA ACGGTCACAT TTTAACACCA TGGGCCATTA 240
TTTGTTAAGC TGACATCTTT AACTCCATTG TACTATGTAT TGTGACTTCG ACCAATTAAA 300
TTGTTTTTTG AGAGTCAATT TAATTGCTGT GTTCATTCTG TTGTGTTATA TATCTGTGGA 360
CAATGTTATG GAACAATACA TATTTCATAA ACATTTGATC ATTTACCTAT GAATTTCACC 420
AGAAATAGAC GTGGTTTGAC AACTACTTCG CGTAGAAGTA GTTATAGGTA TCCTTTGTCT 480
AGACGATCGT ATGGTGTGAA AAGAAACAAT GGTAGACGTG GGCCGAATAT TCCAAACAAG 540
GCCCATGAAG ATGGAAAGAT GTCAGGTCAA TGCATACATG AAAATCAGTT TGGTCCAGAA 600
TTTGTTATGG CCCATAATTC ACCCATTTCA ACTTTCATTA CATACCCTAC TCTGGGTAAA 660
ACTGACTCCA GTCGATCAAG GTCGTATATT AAGTTAAAAC GTCTACGTTT TAAGGGAACT 720
GTGAAAGTTG AACGTCTTAC CCCGAATGTT AACATGGACG GGCTACATCC AAAGATAGAA 780
GGAGTCTTTT CAATGGTGGT TGTGGTCGAT CGAAAACCAC ATTTAAGCCC GTCCGGATGT 840
CTCCACACAT TCGATGAATT GTTTGGTGCC AGGATACACA GTCATGGAAA TTTAGCTATT 900
ACTGCCTCGT TGAAAGACCG ATTCTATATA CGCCATGTGT TTAAACGCGT GATGTCTGTC 960
GAGAAAGATA GCACGATGGT GGATATTGAA GGGACGACAA CATTATCTAA TAGGCGTTTT 1020
AATTGTTGGG CTAGTTTTAG AGATTTTGAA CGTGATTCAT GTAACGGTGT TTATGCCAAC 1080
ATTAGCAAAA ACGCCTTGTT AATTTACTAT TGCTGGATGT CAGATAATGT GTCTAAGGCA 1140
TCTACATTTG TATCATTTGA TCTGGAATAT ATTGGTTGAG TGATAAATTG TGTTTAAACC 1200
TGTCAAGATT TTTATTAAAC GTCAAATGAT AATTTGCACA CGCTATTCAA AAATGCACAA 1260
AGACATTATT TATTTAAGAC ACTTGGGCTC GGAAGGAATA CAGTTGCTTT TAATACATTC 1320
TTGGGCCGTT GTTCTCACCA GCTCGTTCAA TTGGGCCATT GATATTGTTA TACTGGACTC 1380
TGCCCTCCTT GCACCTACTA TTGATGCTGA TTCACCCGGG TCTAAGACAC TGCTGTCTAA 1440
TCTATGCAGT TGTCTATATG GATGGGTTGT GTTCTCAACC TCAGAGTCCG CATCGGTTGA 1500
ACTGACTCCT ATGGTGCTTC TTGTGGCCCA TGTTTCTCCC GGTTTGATCG TTATTGGGCT 1560
GTGTATGCCA TATCTCGTTG TGGATGCGGA CTTGATCAAT TTCCGTTCAT AATGACCATA 1620

```
TCCCACATGA GAAAAATCGA CATCGCTCTC CTTGAACTGT TTCGATAATA TTTTAACTGT 1680
TGGAGCCCTG AAAGGGATAT CAACCGAATG CTTTGCTGTG GAGAGCTTTA ACTTCCCTTT 1740
GAATTTCGCG AAATGTGTTC TTTGATGGAC ATTTGTGTCG CACACTCTAT AGTATAATTT 1800
CCAGGGAATT GGGTCTTTGA GCGAAAAAAA AGACGACGAG AAATAGTGGA GATCTATGTT 1860
ACATCTAATG GGAAACGTCC ATGACGCTTG CAATGACTCA TTGTCCGTCA TTCTCTTGTC 1920
ATGAATCTCC ACTATTACTG ATCCTGTTGC GTTGATAGGA ACCTGTTGTC TGTACTCAAT 1980
CACACAATGG TCGATTTTCA TACAACTACG ACTGATTCTA GCAGTTAACT GAGACGCTGT 2040
GGACGGAAAT TGCAATACTA TCTCAGTTAG GTCATGAGAA AGTTGATATT CATCACGCTG 2100
AGACTCTACG TAATTGAAAA CTGCGGGTGG ATAAGCTAAC TGAGACTCCA TAGAACACAA 2160
TAATAATAAA TAAAAATGGC CGCGCAGCGG AATGGCAACG ACGAACTGAT CAAGGAACAA 2220
GGAAATGTTT TTATTGCTGT GAAAACAGCA GCTGACTAAA GGAAATAAGA AGTTGCTTTC 2280
TTATTGAATA TGTGATTCTT GTTTGGTTTA AATAGACATG ACCATGATCG TACAACCTAA 2340
TTATTAAATA TTATGTCCTC ACATATTCCA TTCATCTTTT ATTTACAAAG AAAAACAAAT 2400
AAAACAATGA TATTGTTAAG AAGTCTGGCA GTGGCATGTT TGTAAATAAA ACGATGTCAC 2460
CAATTGAGCT CGAGCTCCGC TCTAAAACTC TATAGCAATC GGTGAATTGG TGACAATATA 2520
TACTAGAACC CTCTATAGAA CTTCTAATCC TGTTCGCACA CGTGGCGGCC ATCCGCTATA 2580
ATATT                                                             2585
```

CAPÍTULO 2

Comprehensive analysis of the endoplasmic reticulum stress response in the soybean genome: conserved and plant-specific features

Compiled in a paper to be submitted to *BMC Genomics*

Comprehensive analysis of the endoplasmic reticulum stress response in the soybean genome: conserved and plant-specific features

Priscila Alves da Silva[†], José Cleydson F. Silva[†], Hanna D. N. Caetano, Joao Paulo B. Machado, Janaína R. Alves , Giselle C. Mendes, Pedro A.B. Reis, Otavio J.B. Brustolini, Maximiller D.-B. L. Costa, Elizabeth P. B. Fontes*

National Institute of Science and Technology in Plant-Pest Interactions and Departamento de Bioquímica e Biologia Molecular/Bioagro, Universidade Federal de Viçosa, 36570.000, Viçosa, MG, Brazil

*Corresponding author

PAS: priscila.a.alves@ufv.br

JCFS: cleydson@ufv.br

HDNC: hanna.durso@gmail.com

JPBM: jpmachado_bqi@yahoo.com.br

JRA: janabqi@gmail.com

GCM: gisellebio_1@hotmail.com

PABR: pedrobqi@gmail.com

OJBB: otavio.brustolini@ufv.br

MDBLC: maximiller@ufv.br

EPBF: bbfontes@ufv.br

ABSTRACT

Background

The endoplasmic reticulum (ER) is a key signaling organelle involved in the activation of cellular stress responses in eukaryotic cells. Despite the relevance of the ER stress response as an integrator of multiple stress signals into an adaptive response, knowledge about these ER-mediated cytoprotective pathways in soybean (*Glycine max)* is lacking. Here, we searched for genes involved in the highly conserved eukaryotic unfolded protein response (UPR) and ER stress-induced plant-specific cell death signaling pathways in the soybean genome.

Results

As a result of this search, we have provided a complete profile of soybean UPR genes with significant predicted protein similarities to *A. thaliana* UPR-associated proteins. Both arms of the plant UPR were further examined functionally, and evidence is presented that the soybean counterparts are true orthologs of previously characterized UPR transducers in Arabidopsis. The bZIP17/bZI28 orthologs (GmbZIP37 and GmbZIP38) and ZIP60 ortholog (GmbZIP68) from soybean were found to have similar structural organizations as their Arabidopsis counterparts and were induced by ER stress. The truncated forms of GmbZIP37, GmbZIP38 and GmbZIP68 harboring an N-terminal bZIP domain were localized to the nucleus, where they were shown to activate an ERSE- and UPRE-containing BiP promoter. Furthermore, the transcript of the putative substrate of GmIREs, GmbZIP68, was found to harbor a canonical site for IRE1 endonuclease activity and was efficiently spliced under ER stress conditions. Collectively, these data support the notion that the bipartite module of the plant UPR is functionally conserved in soybean. In a reverse approach, we also examined the Arabidopsis genome for components of a previously characterized ER stress-induced cell death signaling response in soybean. With the exception of GmERD15, which apparently does not possess an Arabidopsis ortholog, the Arabidopsis genome harbors conserved GmNRP, GmNAC81, GmNAC30 and GmVPE sequences that share significant structural and sequence similarities with their soybean counterparts. These results suggest that the NRP/GmNAC81+GmNAC30/VPE regulatory circuit may transduce cell death signals in plant species other than soybean.

Conclusions

Our *in silico* analyses, along with current and previous functional data, permitted generation of a comprehensive overview of the ER stress response in soybean as a framework for functional prediction of ER stress signaling components and their possible connections with multiple stress responses.

Keywords: *Glycine max*, unfolded protein response, UPR, programmed cell death, PCD, ER stress**,** UPR transducers, soybean

BACKGROUND

The endoplasmic reticulum (ER) is a highly dynamic organelle that is involved in major cellular functions, such as protein synthesis, the folding and processing of newly synthesized secretory proteins, protein quality control and the maintenance of Ca2+ homeostasis. Due to the tight regulation of ER homeostasis, this organelle is also involved in the activation of cellular stress responses [1-3]. Under normal conditions, the rate of protein synthesis and loading into the ER lumen is balanced with the protein processing capacity of this organelle. Disruption of the equilibrium between protein loading into the ER lumen and the processing and folding capacities promotes the development of a condition known as ER stress. The perturbation of ER homeostasis caused by ER stress often promotes the accumulation of unfolded proteins in the lumen, which triggers a cytoprotective signaling pathway referred to as the unfolded protein response (UPR) [4-6]. Activation of the UPR allows protein loading into the ER lumen to be balanced with the processing and folding capacities under stress conditions. In mammalian cells, the UPR operates as a tripartite module, and the ER stress signal is transduced through ER membrane receptors protein kinase-like ER kinase (PERK), inositol-requiring transmembrane kinase and endonuclease 1α (IRE1) and activation of transcription factor 6 (ATF6) [6]. The restoration of ER homeostasis by the UPR is achieved by (i) transiently slowing down protein synthesis through PERK activation, (ii) up-regulating the expression of ER folding functions and (iii) inducing ER-associated protein degradation (ERAD)-related quality control mechanisms that ensure for the disposal of unfolded proteins through activation of ATF6- and IRE1-mediated UPR arms. However, if the stress persists and UPR fails to restore ER homeostasis, a cell death signal is activated in an attempt to dispose of abnormal cells [1, 6, 7]

In plants, the UPR arms, which are mediated by IRE1 homologs and ATF6-related receptors, have been extensively characterized in Arabidopsis and to some extent in rice and maize [for reviews, see 8, 9]. Apparently, the PERK-like transducing arm of the mammalian UPR is absent from plant cells and thus, the ER stress-induced attenuation of protein synthesis is activated by an as-yet-unknown mechanism. Therefore, upon disruption of ER homeostasis, plant cells activate at

least two branches of the UPR through IRE1-like and ATF6-like transducers, resulting in the up-regulation of ER-resident molecular chaperones and activation of the ER-associated protein degradation system. Recently, a plasma membrane-associated member of the plant-specific NAC domain-containing TF family, AtNAC62, has been demonstrated to undergo cross-talk with ER stress signaling pathways to activate UPR-induced promoters, highlighting a unique aspect of this highly conserved UPR response in plants [10].

Plant IRE1 homologs contain an IRE-like receptor configuration with a stress sensor luminal domain at the N-terminus, a transmembrane segment, and C-terminal kinase and ribonuclease domains. Two IRE1 homologs have been found in Arabidopsis (AtIRE1a and AtIRE1b) and maize and one has been identified in rice (OsIRE1) [11-13]. The substrate for Arabidopsis IRE1 endonuclease activity is the transcript of the ER membrane-associated TF bZIP60, which harbors a hairpin-like structure as the IRE1-specific site and encodes an N-terminal bZIP domain followed by a transmembrane segment [14-16]. In response to ER stress, the endonuclease activity of IRE1 mediates the splicing of bZIP60 mRNA to generate an alternatively spliced transcript that lacks transmembrane domain-encoding sequences. This splicing leads to the synthesis of a soluble and functional bZIP transfactor that can be translocated to the nucleus, where it activates ER stress-inducible promoters. In rice, the ortholog of the Arabidopsis AtbZIP60, OsbZIP74 (also known as OsbZIP50), is also regulated through the OsIRE1-mediated splicing of its RNA [11]. In maize, ZmbZIP60 transcript splicing leads to the activation of ER stress-inducible promoters [13].

The second branch of the UPR in plants mechanistically resembles the ATF6-mediated transduction of the ER stress signal. The Arabidopsis orthologs of mammalian ATF6 include two ER-localized, membrane-tethered TFs, bZIP28 and bZIP17 [17-19]. In the absence of stress, plant BiP is bound to Arabidopsis ATF6-like bZIP28, which remains in the ER membrane [20]. In response to ER stress, BiP dissociates from bZIP28, allowing it to be redirected to the Golgi, where it is proteolytically processed by S1P/S2P and released from the membrane [21, 22]. The released bZIP domain of this transfactor is then translocated to the nucleus, where it acts in concert with the heterotrimeric NF-Y complex to activate UPR genes [23]. In addition to ER stress, bZIP17 is primarily induced by salt stress, a condition that also promotes its regulated movement to the Golgi and S1P/S2P-mediated cleavage,

thereby releasing its N-terminal TF domain for translocation to the nucleus, where it acts in concert with bZIP60 to activate salt stress-responsive promoters and a fraction of ER stress-induced promoters [24-26]. Heat stress induces the expression, S1P/S2P-mediated processing and nuclear translocation of the bZIP28 TF [27]. Maize ZmbZIP17 has been shown to directly link ER stress with ABA signaling [28], and both bZIP28 and bZIP17 connect ER stress and heat stress with BR signaling [29].

The UPR-mediated activation of bZIP60, bZIP17 and bZIP28 promotes the induction of ER-resident molecular chaperones, such as BiP, ERdj, GRP94, CNX, CRT, peptidylprolyl isomerases (PPIases) and thiol disulfide oxidoreductases (PDI and ERp57), through binding to the promoters of the stress-responsive *cis*-regulatory elements UPRE-I and UPRE-II [30]. bZIP60 also transactivates the NAC103 promoter through interaction with a distinct stress-responsive *cis*-regulatory element, UPRE-III [31]. In turn, the NAC103 TF amplifies the UPR signal by further activating several UPR-related chaperones, including CRT1, CNX, and PDI-5 [31]. Downstream components of the UPR also include components of the ERAD machinery, including homologs of EDEMs (MNS4/5), OS9 (EBS6/OS9), Hrd1, Hrd3/Sel1L (EBS5/Hrd3A) and Derlin-1 (Der) [32-35]. Therefore, under moderate stress conditions, the UPR-mediated induction of ER-resident chaperones and ERAD genes promotes ER quality control processes to re-establish ER homeostasis. However, under prolonged and severe stress, if ER functioning and cell growth cannot be restored, then a cell death program is triggered, presumably to protect the organism from aberrant cells that contain unfolded proteins.

One such plant-specific ER stress-induced cell death response has been recently shown to be mediated by regulated intramembrane proteolysis of the ER membrane-tethered NAC089 TF [36]. In response to ER stress, NAC089 is relocated to the nucleus to control the expression of downstream genes involved in PCD, such as NAC094, MC5 and BAG6. Because the expression of NAC089 is controlled by bZIP28 and bZIP60, during the plant ER stress response, these UPR transducers also elicit pro-death signals, a property that is shared by their mammalian counterparts. A distinct plant-specific ER stress-induced cell death response that integrates an osmotic stress signal into a full PCD response has been reported in soybean and is mediated by the developmental cell death domain (DCD)-containing N-rich proteins DCD/NRP-A and DCD/NRP-B [37]. The expression of DCD/NRP is controlled by

the ER and osmotic stress-induced TF GmERD15, which specifically binds to the DCD/NRP promoters to activate the transcription of these genes [38]. Enhanced DCD/NRP accumulation causes the induction of the plant-specific TFs GmNAC81 and GmNAC30, which interact to fully activate expression of the vacuolar processing enzyme (VPE), a plant-specific executioner of programmed cell death (PCD) that displays caspase-1-like activity [39, 40]. Therefore, GmNAC081, GmNAC030 and VPE are involved in a plant-specific regulatory cascade that integrates osmotic stress- and ER stress-induced PCD.

Comprehensive genome-wide evaluations of ER stress-induced changes in gene expression have provided evidence that the UPR operates in a similar fashion in both soybean and Arabidopsis [41]. Inducers of ER stress, such as tunicamycin and AZC, promote the up-regulation of a class of homologous genes that function in (i) protein folding, (ii) ERAD and (iii) translational regulation. As further evidence for the functioning of the UPR in soybean, the promoters of soybean BiP genes contain functional ER stress *cis*-acting elements (ERSEs), and soybean BiP functions as a regulator of the UPR as it does in Arabidopsis [20, 37, 42, 43]. Nevertheless, genes involved in the ER stress response are poorly characterized in soybean, and except for ER stress-induced NRP-mediated cell death signaling, no other branches of ER stress signaling have been examined at the gene level in this plant. In addition, upstream transducers of the UPR have not been functionally or mechanistically identified in the soybean genome. In this investigation, we conducted a complete survey of upstream, immediate downstream and downstream components of the ER stress response in soybean. Additionally, we examined the possible transducer functions of soybean IRE1 homologs and bZIP28/bZIP17-related receptors. Our *in silico* analyses, along with current and previous functional data, have generated a comprehensive overview of the ER stress response in soybean as a framework for functional prediction of ER stress signaling components and their possible connections with multiple stress responses.

RESULTS AND DISCUSSION

The high conservation of the ER stress response in different plant species, such as Arabidopsis and rice, along with the accurate assembly of the soybean genome sequence [44], allowed for the *in silico* identification of components of different branches of the UPR (Table 1) in addition to those of the plant-specific ER stress-induced cell death response (Table 2). Because the plant UPR is transduced as a bipartite module that converges in an adaptive response, we have presented our data in the following groups to facilitate comprehension: UPR transducers/sensors, UPR immediate downstream components and UPR downstream components (Table 1). The corresponding gene copy numbers in the soybean genome are presented in Tables 1 and 2.

Identification of transducers/sensors and immediate downstream components of the UPR

Previously characterized Arabidopsis UPR genes were used as prototypes for the identification of the soybean orthologs and the *in silico* assembly of the UPR in soybean. Using eggNOG v4.0 software, the UPR bZIP transducers bZIP17 and bZIP28 were grouped into the virNOG01396 group, which was comprised of the three genes encoding bZIP17, bZIP28 and bZIP49 (Additional file 1). A search for the bZIP17 and bZIP28 prototypes in eggNOG v4.0 against the Williams 82 v1.1 whole-genome sequence [44] revealed two predicted soybean orthologs (Glyma.03G123200 and Glyma.19G126800, annotated with Phytozome Glyma v.10.1.p, Wm82.a2.v1.1) as the soybean representatives in the virNOG01396 group. A BLASTp search revealed that both of the soybean bZIP gene orthologs were more closely related to bZIP17 (At2G40950). Glyma.03G123200 (GmbZIP38) displayed 60.66% similarity and 48% identity to bZIP17 with 96% protein sequence coverage, and Glyma.19G126800 (GmbZIP37) was 61.32% similar and 47.68% identical to bZIP17 with 94% coverage. The use of the bZIP28 amino acid sequence for comparison resulted in decreases in the similarity and identity of Glyma.03G123200 (GmbZIP38) to 55.99% and 42.11%, respectively, with 80% coverage, whereas Glyma.19G126800 (GmbZIP37) displayed 55.47% similarity and 41.49% identity with 81% coverage (Additional file 1). This level of sequence conservation did not

allow us to distinguish between the bZIP17 and bZIP28 soybean orthologs hence, both GmbZIP37 and GmbZIP38 were assigned as bZIP17/28 orthologs (Table 1).

The immediate downstream components of the bZIP-mediated UPR arm, which are involved in the ER stress-induced mobilization and Golgi-mediated processing of bZIP28 and bZIP17, were also analyzed with eggNOG v4.0. These components included site-1 protease (S1P), a soluble luminal protease, site-2 protease (S2P), a membrane-associated metalloprotease, SAR1, a small GTPase involved in the formation of prebudding complexes for COPII-mediated relocation of cargo from the ER to the Golgi, and SEC12, a COPII vesicle element [8, 45-47]. The copy numbers of the soybean orthologs are shown in Table 1, and the *e-value* showed a high level of conservation of homologous regions between ortholog pairs (Additional file 1).

The Arabidopsis genome contains three copies of the IRE genes, but only IRE1a (At2G17520) and IRE1b (At5G24360) encode full-length proteins (12, 14, 15]. Our *in silico* analysis recovered IRE1a and IRE1b and clustered them into the virNOG09069 group, which encompassed four predicted soybean IRE orthologs. A BLASTp search revealed that the Glyma.01G157800 (GmIRE1a), Glyma.09G197000 (GmIRE1d) and Glyma.11G087200 (GmIRE1c) predicted proteins were the most similar to Arabidopsis IRE1a (80% similarity and 68% identity, but different levels of sequence coverage), whereas Glyma.16G111800 (GmIRE1b) was the most similar to IRE1b (60.67% similarity and 43.14% identity with 94% coverage). The *e-value* showed the high conservation of homologous regions among the orthologous proteins.

A striking feature of the soybean genome is the retention of extended blocks of duplicated genes [44]. Approximately 75% of the 46,430 high-confidence genes predicted to be present in the soybean genome exist as paralogs, and 25% have reverted to singletons [44]. Phylogenetic analysis of the Arabidopsis and soybean IRE orthologs belonging to the virNOG09069 group (Additional file 1) showed that the four soybean IRE paralogs were clustered in pairs, consistent with duplication events (Additional file 2).

bZIP60 is an immediate downstream component of the IRE arm of the UPR, and its transcript has been shown to serve as an IRE substrate [14, 15]. A search of eggNOG v4.0 for the AtbZIP60 sequence against the soybean genome identified just one soybean ortholog, Glyma.02G161100 (GmbZIP68), which was placed into the

euNOG19243 group with significant protein similarity to the AtbZIP60 prototype (Additional file 1). Phylogenetic analysis of soybean and Arabidopsis orthologs of the UPR membrane-tethered bZIP transfactors belonging to the virNOG01396 and virNOG09069 groups confirmed that GmbZIP68 was the most closely related to AtbZIP60 because they were clustered together and separate from the virNOG01396 group of orthologs (Additional file 3). Consistent with a duplication event, soybean GmbZIP37 and GmbZIP38 were clustered together as paralogs, but they were more closely related to AtbZIP17, confirming the eggNOG data (Additional file 1). Predicted protein similarities between soybean and rice bZIP28/17-like genes were also determined (Additional file 3). The orthologous genes in soybean that were the most similar to bZIP17 were also the best matches in rice. The conservation of homologs of these bZIP17-like genes in other species is strongly suggestive of their functional importance and identities.

Recently, an ER stress-induced plant-specific NAC TF, NAC103 (At5g64060), has been shown to be regulated by a functional bZIP60 through the newly identified UPRE-III (TCATCG) on the NAC103 promoter [31]. NAC103 in turn amplifies the UPR signal by up-regulating ER stress-induced promoters, such as CNX and CRT. Using the NAC103 amino acid sequence as a template, we identified four orthologs in the soybean genome (Table 1). eggNOG v4.0 software grouped the NAC103 paralogs ANAC082 and ANAC103 together with the soybean orthologs Glyma.04G213300, Glyma.05G191300, Glyma.08G156500 and Glyma.06G152900 in the virNOG18312 group (Additional file 1). Among them, the Glyma.04G213300 predicted protein, also designated as GmNAC020, displayed the highest sequence similarity to NAC103, whereas Glyma.05G191300 (GmNAC028), Glyma.08G156500 (GmNAC058) and Glyma.06G152900 (GmNAC037) were more similar to ANAC082. As stress-responsive genes, GmNAC020 and GmNAC037 have been shown to be up-regulated by moderate water deficit, whereas GmNAC028 and GmNAC058 are up-regulated by persistent water deficit conditions [48]. The deduced protein sequences of all four NAC103 orthologs from soybean were found to contain a highly conserved NAC domain at the N-terminus that was divided into five NAC subdomains (A–E) of conserved blocks. The presence of an ER stress-responsive element controlled by bZIP60, pUPRE-III, was identified on the GmNAC028 promoter.

The bZIP- and IRE-mediated arms of the plant UPR are functionally conserved in soybean

The AtbZIP17 and AtbZIP28 TFs are proteolytically activated by inducers of ER stress, such as tunicamycin and DTT, and by adverse environmental conditions, such as heat and salinity [19, 26]. As a consequence, the bZIP domain is released from the membrane and enters into the nucleus, where it regulates the expression of UPR-responsive genes controlled by pERSE (CCAAT-N10-CACG)-, pUPRE (ATTGGTCCACGGTCCATC)-, pUPRE-I (TGACGT-GR)-, pUPRE-II (GATGACGCGTAC)- and/or pUPRE-III (TCATCG or CGATGA)-containing promoters [23, 31, 35, 42, 49, 50-53].

The functions of soybean GmbZIP38 and GmbZIP37 as bZIP17/28-like UPR transducers were examined using several different approaches. We first analyzed the expression profiles of GmbZIP37 and GmbZIP38 in response to stress conditions known to promote accumulation of unfolded proteins in the ER and to induce AtbZIP17/28 expression, such as ER stress and salt stress. The treatment of soybean seedlings with the salt stress inducer NaCl and the ER stress inducer tunicamycin (which blocks protein glycosylation in this organelle) induced accumulation of the GmbZIP38 (Figure 1A and 1B) and GmbZIP37 transcripts with similar kinetics (Figure 1A and 1B). Controls for the effectiveness of the salt and ER stress treatments, such as GmNAC035 [54] and BiP (soyBIPD) [55], were also included in the assay (Figure 1). GmbZIP38 and GmbZIP37 display patterns of expression similar to those of AtbZIP17 and AtbZIP28.

To further examine whether GmZIP37 and GmbZIP38 function in the soybean UPR, we searched for bZIP17/28 functional domains in the predicted primary structures of GmbZIP38 and GmbZIP37. Several conserved motifs were found in the GmbZIP38 and GmbZIP37 sequences at corresponding positions in the AtZIP orthologs (Additional files 4 and 5)Figs. S1 and S4). These motifs included a bZIP domain at the N-terminal cytosolic-facing region of the predicted proteins, followed by a transmembrane segment and a canonical S1P site (RXXL or RXLX) [19, 30] at the luminal C-terminus (Additional file 5, boxed sequence, RRTL). Based on the mechanistic model of bZIP28 activation and the conserved motifs present in GmZIPs, one may predict that the proteolytic release of GmZIP38 and GmbZIP37 from the ER membrane would render the nuclear-localized bZIP domain functional for regulation of ER stress-induced promoters. To clarify this process, we prepared

truncated versions of GmbZIP38 (bZIP38$^{1-434}$) and GmbZIP37 (bZIP37$^{1-406}$) harboring an N-terminal bZIP domain without the transmembrane segment that was fused to YFP, under the control of the 35S promoter (Additional file 4). The truncated YFP-bZIP38$^{1-434}$- and YFP-bZIP37$^{1-406}$ fusion constructs co-localized with the nuclear marker AtWWP1 fused to mCherry in the nuclei of *N. tabacum* epidermal cells when they were transiently co-expressed in leaves (Figure 2A and 2B, merged). To provide further evidence that GmbZIP37 and GmbZIP38 are functionally linked to the UPR signaling pathway, we examined whether the truncated bZIP domains directly target ER stress *cis*-regulatory element-containing promoters. We performed *β*-glucuronidase (GUS) transactivation assay using the −2200pbip9-gus tobacco transgenic line stably transformed with a *β-GUS* reporter gene under control of the soyBiPD promoter [42]. The soyBiPD promoter harbors repeated ERSEs (with the coordinates −552 to −534, −280 to −260, −219 to −201 and −193 to −175) and a UPRE-I (with the coordinates −185 to −175), which have been previously shown to function as ER stress-responsive elements [42]. We also assessed a control transgenic line stably transformed with a promoterless GUS gene (pCambia empty vector). Accumulation of YFP-bZIP38$^{1-434}$ and YFP-bZIP37$^{1-406}$ transcripts in agroinfiltrated −2200pbip9-gus transgenic leaves and in pCambia control leaves was confirmed by qRT-PCR (Figure 3A). The effects of promoter transactivation were assessed by measuring β-galactosidase activity (Figure 3B), as well as by quantifying reporter GUS transcript accumulation (Figure 3C). The bZIP domains bZIP38$^{1-434}$ and bZIP37$^{1-406}$ specifically activated the BiP promoter, enhancing GUS activity and inducing GUS transcript accumulation in the 2200pbip9-gus transgenic leaves compared with the pCambia transgenic leaves. The infiltration of untransformed *Agrobacterium* culture (Gv3101) and the expression of GFP alone in −2200pbip9-gus transgenic leaves did not result in targeting of the BiP promoter. Collectively, these results implicate GmbZIP37 and GmbZIP38 as true orthologs of Arabidopsis bZIP28 and bZIP17 and suggest that the bZIP28-mediated arm of the UPR is mechanistically conserved in soybean.

As the second arm of the plant UPR signaling pathway, upon activation, the dual-functioning protein kinase/ribonuclease IRE1 initiates transduction of the ER stress signal by splicing the bZIP60 transcript. The spliced bZIP60 transcript encodes a truncated version of the protein that lacks a transmembrane domain and C terminus therefore, the N-terminal bZIP domain is capable of translocating to the nucleus to

activate ER stress-responsive genes. In Arabidopsis, bZIP60 is activated and induced by ER stressors and diverse environmental stress conditions that promote accumulation of unfolded proteins in the ER [14, 49]. We found that the bZIP60 ortholog from soybean, GmbZIP68 (Glyma.02G161100), was also induced by the ER stressor tunicamycin (Figure 4). To examine whether the GmbZIP68 transcript undergoes IRE1-mediated unconventional splicing, we first searched for potential hairpin-like IRE1-specific sites in the GmbZIP68 transcript using RNA structure prediction software (Mfold v2.3). The predicted form of the GmbZIP68 transcript with the lowest free energy is presented in Additional file 6, from which we selected a pair of adjacent hairpin loops, with three conserved bases in each loop (Figure 5A - B). The selected double-hairpin structure resembles the splicing site in the bZIP60 transcript, which is specifically cleaved by IRE1 at a conserved sequence (CUG↓CUG) in each loop [14-16]. Based on this bZIP68 twin stem-loop structure with a conserved splice sequence for IRE1 in each loop, we predicted that GmIRE1-mediated alternative splicing of GmZIP68 would remove a 23-nucleotide segment from that site, causing a translational frameshift in the spliced RNA to precisely delete the transmembrane domain, rendering a soluble, functional protein (Additional file 7). To address this possibility, we designed two sets of primers that were each specific for spliced (bZIP68s) or unspliced (bZIP68u) GmZIP68 mRNA (Additional file 7). Primer specificity was confirmed by RT-PCR using cDNA prepared from tunicamycin-treated and untreated soybean seedlings (Additional file 8A). The primers specific for unspliced bZIP68u amplified fragments from both tunicamycin-treated and untreated seedling RNA (lanes 1, 4 and 6), whereas those for spliced bZIP68s amplified a fragment from tunicamycin-treated seedling RNA (lanes 3 and 5) but failed to amplify it from untreated seedling RNA (lane 2), which is consistent with an ER stress-induced splicing event in the target RNA. To detect the removal of the predicted 23b segment of RNA (Figure 5A and 5B), the RT-PCR products were separated with a 15% (w/v) polyacrylamide gel, and RT-PCR was performed using the two sets of primers in the same reaction (Figure 5C). RNA from untreated soybean seedlings produced an RT-PCR product with a single band on the polyacrylamide gel (lane 1), whereas that extracted from soybean seedlings treated with tunicamycin for 8 h and 24 h generated RT-PCR products with two bands, confirming the ER stress-mediated splicing of GmbZIP68. As a positive control for the ER stress-induced splicing assay, RT-PCR using tunicamycin-treated

Arabidopsis RNA with bZIP60u- and bZIP60s-specific primers resulted in the expected double band on a polyacrylamide gel (Additional file 8B).

We also determined the transcription-regulating activity of GmbZIP68 by performing GUS transactivation assays using −2200pbip9-gus tobacco transgenic leaves. We first transiently expressed an N-terminal truncated version of GmbZIP68 (up to aa position 209) fused to YFP by agroinfiltrating a 35S::bZIP68$^{1-209}$-YFP construct into *N. benthamiana* leaves and examining its subcellular localization by confocal microscopy (Additional file 4). As expected for a truncated bZIP protein with no transmembrane segment, fluorescence of the YFP-GmbZIP68$^{1-209}$ fusion protein was concentrated in the nuclei of agroinfiltrated *N. benthamiana* leaves (Figure 2C), co-localizing with the nuclear marker AtWWP1-mCherry (merged). For GUS transactivation assay, *Agrobacterium* carrying a 35S::YFP-GmbZIP68$^{1-209}$ construct or a 35S::GFP negative control was infiltrated into 2200pbip9-gus transgenic leaves and pCambia transgenic leaves. Expression of YFP-GmbZIP68$^{1-209}$, but not that of GFP, activated the BiP promoter, as determined by increases in β-galactosidase activity (Figure 3B) and transcript levels (Figure 3C) in the 2200pbip9-gus transgenic leaves compared with the controls. Collectively, these results indicate that GmbZIP68 is a true ortholog of bZIP60 from Arabidopsis and that the IRE-mediated splicing arm of the UPR is functionally conserved in soybean.

Analysis of UPR downstream components in the soybean genome

To restore ER homeostasis under stress conditions, the plant UPR signaling pathway elicits the up-regulation of stress-specific responses, including increases in protein folding and degradation in the ER. The protein folding capacity of the ER depends on the repertoire of resident molecular chaperones, which has been extensively characterized in Arabidopsis [9, 30]. Therefore, we used the known chaperones from Arabidopsis as the prototypes to identify soybean orthologs though searches of eggNOG. We also searched for typical domains in ER-resident proteins, such as N-terminal peptide signals and C-terminal ER retention signals, as additional criteria to identify soybean orthologs. High degrees of sequence identity/similarity and highly significant e-values were consistently observed between the orthologous pairs of UPR downstream components.

BiP, the most abundant chaperone in the ER, belongs to the heat shock protein 70 kDa (HSP70) family and has been extensively characterized in different plant species, such as tobacco, soybean and Arabidopsis [42, 55-60]. In addition to its molecular chaperone activity, plant BiP also functions in regulating signaling events related to ER stress, and it displays protective functions under distinct stress conditions, including the attenuation of ER stress [37, 61, 62], the promotion of drought tolerance in transgenic soybean (*Glycine max*) and tobacco (*Nicotiana tabacum*) plants [61, 63], the activation of plant innate immunity [64] and the attenuation ER stress- and osmotic stress-induced cell death in soybean [65]. In general, plant BiP is represented by multiple copies (Table 1). A search of eggNOG v4.0 using AtBiP1 as the prototype resulted in the placement of BiP1 and BiP2 from Arabidopsis into the virNOG09258 group together with the previously described soybean BiPs (soyBiPA, soyBiPB/soyBiPD and soyBiPC) and a new additional soybean BiP, gene model Glyma.05G219600 (Additional file 1). AtBiP3 was grouped separately as a virNOG29237 representative, with no closely related homolog in soybean. Phylogenetic analysis of HSP70 members from Arabidopsis and soybean revealed that the BiP proteins were clustered together in a distinct clade, with AtBiP3 placed separate from the others, representing the most distant member of the family (Additional file 9).

Calnexin (CXN) and calreticulin (CRT) represent the major protein folding machinery of the ER, and they specifically bind glycoproteins that carry monoglucosylated N-linked glycans [9]. Calreticulin is a soluble protein in the ER lumen, whereas calnexin is a type 1 membrane protein. In Arabidopsis, three CRT isoforms and two CNX isoforms have been described [34, 66, 67]. A search of eggNOG for the At1G08450 (CRT3) protein sequence against the soybean genome resulted in the clustering of four soybean CRT paralogs into the virNOG02900 group. The other two Arabidopsis CRTs (CRT-1a and CRT-1b) were placed into the virNOG10578 group, together with two soybean CRT orthologs (Additional file 1). The confidence index e-value revealed high conservation of homologous regions between orthologous pairs.

The *A. thaliana* CNX genes were recovered and clustered into five different groups (virNOG06123, virNOG06264, virNOG09352, virNOG13792 and virNOG23307), which included 5 Arabidopsis genes and 11 soybean orthologs (Table 1 Additional file 1). The largest group, virNOG13792, contained the most

well-characterized Arabidopsis CNX gene, CNX1 (At5G61790) [7], and four soybean orthologs. All members of the CNX family displayed remarkably conserved primary structures and the conserved domains of ER-resident proteins. As ER-resident molecular chaperones, both calreticulin and calnexin from soybean are induced by ER stressors, such as tunicamycin and AZC [41].

ER-resident protein disulfide isomerases (PDIs), which are associated with the CNX/CRT system, catalyze disulfide bond formation, which plays relevant roles in the folding and stabilization of tertiary and quaternary protein structures [7, 9, 68]. PDIs are multi-domain proteins that belong to the thioredoxin (TRX) superfamily and hence harbor at least one TRX domain [69]. The Arabidopsis genome encodes 13 PDIs, but only 9 possess known ER retention signals and have been implicated in protein folding [34, 68-74]. In the soybean genome, 22 PDI paralogs have been previously identified [70]. Our analysis did not result in the selection of the gene model Glyma12g16310 (Phytozome v9.1 as in 70) as a PDI instead, Glyma.14G152000 (Phytozome v10.1) was included as a new additional PDI paralog in soybean (Additional file 1).

The Arabidopsis PDI paralogs At2G47470, At3G20560 and At4G27080 do not have known ER retention signals [68]. At2G47470 and four highly conserved soybean orthologs (more than 82% similarity, 70% identity and 86% sequence coverage Additional file 1) were clustered into the virNOG09353 group. The other two PDI paralogs lacking known ER retention signals, At3G20560 and At4G27080, were placed with three soybean orthologs into the virNOG04036 group. The members of this latter group displayed high degrees of sequence conservation with more than 85% similarity, 71% identity, 97% coverage and significant e-values. The soybean PDI orthologs also did not harbor known ER retention signals.

The remaining Arabidopsis PDIs with ER retention signals and 15 soybean orthologs were distributed into seven distinct eggNOG v4.0-generated groups that were significantly conserved (Additional file 1). Phylogenetic analysis results recapitulated the eggNOG data (Additional file 10). PIN1, ERdj and GRP94 orthologs are also represented in the soybean genome by small gene families (Table 1, Additional file 1)

We also identified predicted soybean orthologs involved in glycoprotein folding, such as oligosaccharyltransferase (OST), glucosidase I (Glc-I), glucosidase II (Glc-II) and UDP-glucose:glycoprotein glucosyltransferase (UGGT), using the

Arabidopsis homologs as prototypes (Additional file 1) [75-78]. Remarkable sequence conservation among all orthologs in this category was supported by their significant e-values, and their high levels of similarity and identity suggest that they are functional analogs.

We also examined the components of ERAD in the soybean genome as downstream components of the UPR. The degradation of unfolded proteins by the ERAD system is crucial for the re-establishment of ER homeostasis under stress conditions and involves the following four steps: (i) recognition, (ii) ubiquitination, (iii) retrotranslocation and (iv) protein degradation [9, 79]. In yeast and mammals, the ERAD pathway has been intensively characterized, and this information has been used to identify orthologs in the Arabidopsis genome [34, 80]. The profile of plant ERAD components was extended in this current investigation to include predicted orthologs in soybean. Usa1-like, Cue1-like and OTU1-like proteins were not found in Arabidopsis or soybean. HRD3B–like, PUX6-like, PUX11, PUX12, PUX14 and PUX15 proteins were not detected in the soybean genome. All other ERAD components were represented by at least two related copies in the soybean genome (Table 1). In general, sequence comparison analyses revealed that the predicted ERAD-associated orthologous pairs in Arabidopsis and soybean shared significant amino acid sequence conservation (e-values<10E-10, sequence similarities and identities of higher than 50%, with protein sequence coverage of greater than 70%).

Recently the N-glycan ERAD pathway, which monitors the correct glycosylation of proteins and targets improperly folded glycoproteins for degradation, has been shown to be highly conserved in plants [81-85]. Terminally unfolded glycoproteins are removed from the CNX/CRT folding system through the sequential hydrolysis of two α1,2-mannose residues, as mediated by the ER-resident α1,2-mannosidase MNS3 in Arabidopsis (MNS1 in yeast and mammals and two soybean predicted orthologs) and MNS4/MNS5, which corresponds to homologous to mannosidase 1 (Htm1) in yeast, ER-degradation enhancing a-mannosidase-like protein (EDEM) in mammals and three predicted orthologs in soybean (Additional file 1). The removal of the α1,2-mannose residue exposes a α1,6-mannose residue, which is a targeting signal for the ER-resident lectins EBS6 (Os9 in mammals and Yos9 in yeast) [34, 80, 84, 85] and EBS5/HRD3A (HMG-CoA reductase degradation 3 (Hrd3) in yeast and suppressor of lin-12-like (Sel1L) in mammals) [86]. EBS6 and EBS5 recruit and targets unfolded proteins to the ER membrane-tethered ERAD

complex for ubiquitination and retrotranslocation. The Arabidopsis genome has two Hrd3/Sel1L homologs, designated as AtSel1A (EBS5 or HRD3A) and AtSel1B (HRD3B, an apparent pseudogene) [86], and an Os9/Yos9 homolog, AtOs9 (EBS6) [80, 85], whereas in the soybean genome, two representatives in each class with significant sequence identities to their Arabidopsis counterparts were identified (Table 1 and Additional file 1).

The central component of the ERAD complex is a cytosolic-facing ER membrane-associated E3 RING finger-type ubiquitin ligase responsible for the ubiquitination of ERAD substrates and for connection of a series of luminal and cytosolic adapters [79]. The HMG-CoA reductase degradation (Hrd1) protein is specific for ERAD L/M substrates (from membrane or lumen) and degradation of alpha2 (Doa10) ERAD C substrates (from cytosol). The Arabidopsis genome encodes two Hrd1 orthologs (AtHrd1A and AtHrd1B) [86] and two Doa10 orthologs (Doa10A and Doa10B) [18, 79, 87]. Likewise, in the soybean genome, we identified two copies with significant similarities to Hrd1 and Doa10, respectively (Table 1 and Additional file 1). In yeast, Hrd1 E3 ligase-associated proteins include Cue1 (ER anchor protein), UBC6 (membrane-anchored E2), U1-Snp1 associating-1 (Usa1 HERP in mammals) and degradation in the ER (Der1 Derlin, Der1-like protein in mammals). The proteins Cue1 and UBC6 are also contained in the Doa E3 ligase complex. In Arabidopsis, three UBC6 (UBC32, UBC33 and UBC34) homologs are associated with Doa10 [88-90], whereas in the soybean genome, UBC32 is represented by two copies and UBC34 by two copies. *UBC32* (Ubc6-like E2*)* is induced by salt, drought and ER stress [88]. Neither the Arabidopsis nor the soybean genome encodes the Usa1 or Cue1 gene, but they contain three and four Der1 homologs, respectively [33, 91] (Additional file 1).

Ubiquitinated ERAD substrates are extracted from the ER lumen (ERAD L substrates) or from the ER membrane (M/C substrates) by a trimeric complex of the homohexameric proteins cdc48 (p97 or valosin-containing protein in mammals), Ufd1 and Npl4 (each harboring an ubiquitin-binding domain) [92]. The cdc48/Ufd1/Npl4 complex is recruited by the E3 Hrd1/Doa10 E3 complex through Ubx2 (p97/VCP-interacting membrane protein in mammals, VIMP). The resulting polyubiquitinated ERAD substrates are further processed through antagonistic interactions between ufd2 (U-box-containing E4 multiubiquitination enzyme) and ufd3 (WD40 repeat-containing protein) in addition to Otu1 (deubiquitylating

enzyme) and/or by Png1 peptide (cytoplasmic peptide: N-glycanase, PNGase)-mediated deglycosylation [93]. Processed ERAD substrates are directed to the 26S proteasome by Cdc48 and two ubiquitin receptors, Rad23 and Dsk2, for degradation. The Arabidopsis genome encodes three cdc48 homologs (AtCDC48A, AtCDC48B, and AtCDC48C) [94], whereas in the soybean genome, there are 10 predicted proteins with significant similarity to cdc48 (Additional file 1). AtCDC48 is recruited to the ER membrane by UBX domain-containing proteins, which are represented by 15 copies (AtPUXs) in the Arabidopsis genome and interact with AtCDC48A [95, 96]. A total of 17 PUX homologs were identified in the soybean genome (Table 1). While Ufd1, Ufd2, Ufd3, Npl4, Rad23, and Dsk2 are encoded by gene families in Arabidopsis and soybean, PNG is a single-copy gene in the Arabidopsis genome and is represented by two copies in the soybean genome [34]. The high levels of conservation of primary sequences and domain structures among the ERAD components from yeast, mammals and two plant species, along with the findings of functional studies of Arabidopsis [for a review, see 34] and expression analyses of soybean and Arabidopsis that have been conducted [32, 35, 41], support the notion that the ERAD system functions in plants in a similar manner as in mammals and yeast.

Identification of ER stress-induced plasma membrane-associated NAC062 homologs in the soybean genome

Recently, a plasma membrane-tethered member of the NAC family, NAC062, has been shown to integrate UPR signaling through an as-yet-unknown mechanism. ER stress causes the release of the NAC domain from the plasma membrane and its relocation to the nucleus to regulate ER stress-responsive genes [10]. NAC062 expression is controlled by bZIP60. Using eggNOG v4.0 software, GmNAC062 was determined to be a member of the virNOG05505 group, which is comprised of three genes, ANAC062, CBNAC and ANAC091, in addition to three predicted soybean orthologs, GmNAC021, GmNAC036 and GmNAC110 (Table 1 Additional file 1). In addition to displaying significant amino acid sequence similarity, as determined by comparing orthologous pairs, the NAC062 orthologs possess a predicted transmembrane segment and an N-terminal peptide signal that may target them to the plasma membrane. The promoters of the three soybean ortholog genes GmNAC021,

GmNAC036 and GmNAC110 harbor a UPR *cis*-regulatory element, pUPRE-III (TCATCG), which is a bZIP60 binding site [31].

Plant-specific ER stress-induced cell death responses may be conserved in soybean and Arabidopsis

Another plant-specific NAC domain-containing TF, GmNAC089, which is an ER membrane-associated protein, has been shown to play a relevant role in the ER stress response by positively regulating ER stress-induced PCD [36]. ER stress causes relocation of GmNAC089 from the ER membrane to the nucleus, where it induces the expression of PCD-associated genes. A search of eggNOG for GmNAC089 against the soybean genome did not identify any soybean orthologs. Two predicted soybean orthologs, GmNAC103 [97] and an as-yet-unclassified soybean NAC gene (Glyma.12G186900), were identified using BLASTp. This novel full-length NAC predicted protein was recovered from the recently released version of the revised soybean genome, *Glycine max* Wm82.a2.v1, suggesting the existence of two homologous copies of NAC089 in the soybean genome. These soybean orthologs harbor a predicted transmembrane segment and an N-terminal peptide signal that may target them to the ER membrane (Table 2, Additional file 11).

The ER stress- and osmotic stress-induced NRP/DCD-mediated cell death cooperative response, which has been described in soybean, may be the most well-characterized plant-specific ER stress-induced PCD signaling response [37-40]. Consequently, we used soybean genes as prototypes to identify orthologs in Arabidopsis. The GmERD15 (Glyma.14G055200) TF is the most upstream component characterized, and it is induced by osmotic and ER stress to trigger the expression of NRP/DCD genes. A search of eggNOG recovered two paralogs in the soybean genome clustered in the virNOG24368 group and no ortholog in the Arabidopsis genome (Table 2, Additional file 11).

A search of eggNOG for the DCD/NRP-A sequence identified two paralogous copies each of DCD/NRP-A and DCD/NRP-B and a single-copy gene, AtNRP1/At5G42050, in the Arabidopsis genome, with significant amino acid sequence similarities (Additional file 11). These genes were clustered into the virNOG01663 group separately from two soybean paralogs and an Arabidopsis ortholog (AtNRP2/At3G27090) of DCD/NRP-C, which were placed into the virNOG01663 group. Phylogenetic analysis confirmed the separation of DCD/NRP-

C from the DCD/NRP-A-DCD/NRP-B cluster (Additional file 12). Although DCD/NRP-A and DCD/NRP-B have redundant and relevant functions in cell death signaling, it remains to be determined whether DCD/NRP-C also functions in the transduction pathway. The high degree of sequence conservation among NRP orthologs from three plant species (rice genome was included in the analysis) may also implicate functional conservation. Consistent with this hypothesis, AtNRP1 and AtNRP2 display similar expression profiles and subcellular localizations as the soybean orthologs [98].

The execution of the cell death program has been proposed to occur through NRP-mediated induction of the GmNAC081-GmNAC030-VPE module [40]. GmNAC081 and its paralog were placed into the virNOG11218 group together with an Arabidopsis ortholog (ANAC036/At2G17040 Additional file 11). In contrast, GmNAC30 was found to be represented by a small multigene family, with seven copies in the soybean genome, which were placed together with the Arabidopsis ortholog ANAC002/ATAF1 (At1G01720) into the virNOG09836 group. Phylogenetic analysis based on the NAC sequences involved in UPR signaling and the ER stress cell death response confirmed the eggNOG data, further supporting the notion that the ER stress-induced NAC orthologs share conserved unique functions in the plant ER-stress response (Additional file 13).

The VPE family has five representatives in the soybean genome [40]. A search of eggNOG for the Glyma.14G092800 sequence against the Arabidopsis and soybean genomes recovered and resulted in the grouping together of four soybean paralogs and two Arabidopsis orthologs (At2G25940/alphaVPE and At4G32940/gammaVPE) into the virNOG04445 group. These data were confirmed by phylogenetic analysis, which revealed that the four soybean VPEs and alphaVPE and gammaVPE from Arabidopsis formed a unique clade that was separated from the fifth soybean VPE, Glyma01g05135 (Additional file 14). The four most closely related soybean VPEs display similar expression profiles during development and in response to stress [43]. The expression profiles and functions of more distantly related VPEs have not been examined. The high conservation of the components of the ER stress NRP-mediated cell death response between soybean and Arabidopsis suggests that this cell death signaling response may be a general ER stress response in plants rather a specific transduction pathway in soybean.

CONCLUSIONS

The remarkable conservation of specific branches of the ER stress response, such as cytoprotective UPR signaling, among eukaryotes along with the availability of genome sequences have accelerated the characterization of relevant signaling branches that emanate from the ER. In fact, the characterization of UPR signaling in Arabidopsis was based on mammalian groundwork associated with genetic reverse identification and functional studies of downstream components. Despite the relevance of the ER as a key organelle involved in stress adaptive responses, genes involved in the ER stress response in soybean have not been examined to date. Here, we present a complete repertoire of the potential players in the soybean ER stress response, generating a comprehensive panel as a framework for functional predictions.

As the major result of our research, a complete scenario of the ER stress response in soybean is presented in Figure 6. An interactive map of this comprehensive panel of the ER stress response is also available at the address http://inctipp.bioagro.ufv.br/upr/. This tool enables the access of detailed information about the protein families in the soybean database by clicking on the representative gene in the panel. In Figure 6 and in its online version, the normal pathway of secretory proteins as they enter the ER lumen and proceed towards the Golgi is depicted in numbers 1 through 6. Disruption of proper folding results in deviation from this route (number 6) to a protein degradation pathway, shown in numbers 7 through 9. Accumulation of unfolded proteins activates UPR signaling, which functions as a bipartite module. The ATF6-like-mediated arm of the UPR can be followed in numbers 11, 14, 15 and 16, whereas the IRE1-like-mediated arm is presented as route 12. Plant-specific cross-talk between ER stress response pathways and plasma membrane-associated proteins is presented as route 17. We also identified representatives of the plant-specific ER stress-induced cell death response in the soybean genome. In route 20, transduction of an ER stress-induced signal starts with the predicted regulated intramembrane proteolysis of a membrane-tethered NAC domain-containing TF. The mechanism of execution of the cell death program is lacking, with the exception of the observation that the released NAC TF up-regulates cell death-associated genes. The ER stress- and osmotic stress-induced cell death response is initiated in number 22 and culminates with activation of the

expression of VPE (number 24), which is an executioner of plant-specific vacuole collapse-mediated PCD. In soybean, similar to other eukaryotic organisms, ER stress triggers the evolutionarily conservative UPR and also accommodates cross-talk with several other adaptive signaling responses, such as osmotic-stress induced cell death and ER stress-induced PCD.

We provided several lines of evidence that the soybean and Arabidopsis ER stress responses operate similarly. First, in addition to the high conservation of the primary structures of the soybean and Arabidopsis putative orthologs, they share common functional and localization domains that may be associated with their shared biochemical activities and subcellular localizations. In general, ER-resident protein orthologs harbor a signal peptide and an ER retention signal and transmembrane receptors, which undergo regulated intramembrane proteolysis, possess a predicted transmembrane segment and nuclear localization signal. Second, both arms of the UPR were further examined functionally, and evidence is presented that the soybean counterparts are true orthologs of previously characterized UPR transducers in Arabidopsis. The bZIP17/bZI28 orthologs (GmbZIP37 and GmbZIP38) and ZIP60 ortholog (GmbZIP68) from soybean share similar structural organizations as their Arabidopsis counterparts and are induced by ER stress. The truncated forms of GmbZIP37, GmbZIP38 and GmbZIP68 harboring an N-terminal bZIP domain were localized to the nucleus, where they were shown to activate an ERSE- and UPRE-containing BiP promoter. Furthermore, the transcript of the putative substrate of GmIREs, GmbZIP68, harbors a canonical site for IRE1 endonuclease activity and is efficiently spliced under ER stress conditions. The ER stress-induced splicing of bZIP68 causes a translational frameshift in the C-terminal coding region, generating a truncated version of the protein that lacks the transmembrane segment and includes a second nuclear localization signal. These expression and functional analyses of bZIP38, bZIP37 and bZIP68 support the notion that the bipartite module of the plant UPR is functionally conserved in soybean. Finally, in a reverse approach, we showed that the ER stress-induced DCD/NRP-mediated PCD response, which has been previously characterized in soybean, is also conserved in Arabidopsis. With the exception of GmERD15, which apparently does not possess an Arabidopsis ortholog, the Arabidopsis genome harbors conserved GmNRP, GmNAC81, GmNAC30 and GmVPE sequences that share significant structural and sequence similarities with their soybean counterparts. These

components have also been identified in the rice genome, further indicating that plant-specific NRP-mediated cell death signaling is conserved in other plant species. The components of the ER and osmotic stress-induced cell death signaling pathways are also induced by other biotic and abiotic signals, such drought, salt and pathogen incompatible interactions [39, 48, 54]. Therefore, activation of the NRP-mediated PCD response is not specific to ER stress or osmotic stress rather, it is a shared branch of general environmental adaptive pathways.

METHODS

In silico identification of unfolded protein response components in soybean

We first searched in the literature for previously described *Arabidopsis thaliana* UPR genes, including those encoding both upstream receptors (ER stress sensors) and downstream components involved in ERAD and the processing and folding of secretory proteins. These Arabidopsis genes were used as reference genes for the identification of UPR components in soybean (Additional file 1, reference list). The soybean genes involved in ER stress-induced NRP-mediated cell death signaling, a plant-specific ER stress-induced cell death response that has been previously described in soybean, were used as prototypes to search for counterparts in the Arabidopsis genome (Additional file 11, reference list). Using eggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) database v.4.0 (http://eggnog.embl.de) [99], we identified orthologous plant genes from virNOG (Virideplantae NOG, ftp://eggnog.embl.de/eggNOG/4.0/members/virNOG.members.txt.gz)-predicted groups. EggNOG is a frequently updated database of orthologous genes/proteins with the following identification steps: (i) the use of 23 reference genomes from different plant species (ii) a comparison of amino acid sequences with a FASTA algorithm and (iii) the use of a quality filter to remove low-complexity sequences, generating orthologous groups. Then, the alignment scores between orthologous protein pairs are re-calculated by adjusting for the composition of the amino acid substitution matrix without the

low-complexity regions. Each group of orthologous genes is further characterized functionally using annotations in CG, KOG, arCOG databases and functional Gene Ontology categories, in addition to application of a machine learning method commonly known as support vector machine, which uses annotations from KEGG, SMART, Pfam and Gene Ontology as training features [99].

Using a locally developed script, Arabidopsis and soybean orthologous genes were extracted from eggNOG v4 database (Supplementary Tables S1 and S2). Because eggNOG v4 includes the soybean genome assembly *Glycine max v1.1* (http://www.phytozome.net/search.php?show=text&method=Org_Gmax), we also used the soybean genome assembly *Glycine max* Wm82.a2.v1 that was recently released by Phytozome v10.1 (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax) to update the annotations of the version v1.1 genes. Therefore, the annotations of the soybean genes were recovered from Phytozome v10.1 (http://phytozome.jgi.doe.gov/).

The group of *A. thaliana* orthologous genes initially recovered from the virNOG dataset (ftp://eggnog.embl.de/eggNOG/4.0/members/virNOG.members.txt.gz) were complemented by performing searches of the euNOG (ftp://eggnog.embl.de/eggNOG/4.0/members/euNOG.members.txt.gz) and KOG modules implemented in eggNOG v4.0 (Supplementary Tables S1 and S2). The amino acid sequences of orthologous genes from *A. thaliana* and soybean were recovered from TAIR (http://arabidopsis.org/) and Phytozome v10.1 databases, respectively. Pairwise amino acid sequence comparisons between each member of a group were performed using Basic Local Alignment Search Tool (BLAST), with an *e-value* cut off of ≤ 10E-10. This approach led to the identification of soybean orthologs that were more closely related to the Arabidopsis genes based on the criteria of greater identity,

similarity and sequence coverage, which were supported by the *e-value* confidence index.

In silico validation of the AtbZIP17, AtbZIP28 and AtbZIP60 orthologous genes from soybean

For the *in silico* functional characterization of AtbZIP17, AtbZIP28, AtbZIP60, AtNAC089 and AtNAC062 orthologs from soybean, we analyzed the presence of protein domains, the putative locations of the orthologous predicted proteins and the topology of transmembrane helices. For the identification of protein domains, we used PFAM database (http://pfam.xfam.org/) and HMMer tools (hmmer.janelia.org), which are both available in SMART v.7.0 web server (http://smart.embl-heidelberg.de/). The transmembrane helix segments were identified using TMHMM software (http://www.cbs.dtu.dk/services/TMHMM/) [100].

Phylogenetic analyses of the IREs, bZIP17, bZIP28 and bZIP6*0 genes*

Initially, we constructed a dataset that included the IRE genes from soybean (Glyma.01G157800, Glyma.09G197000, Glyma.11G087200, and Glyma.16G111800) and the Arabidopsis orthologous genes (AT2G17520 and AT5G24360), all of which belonged to the virNOG09069 group (Additional file 1). A second dataset contained the soybean bZIP genes (Glyma.03G123200 and Glyma.19G126800, Glyma.02G161100) and the potential orthologous genes from *A. thaliana* (AT2G40950, AT3G10800, AT3G56660, and AT1G42990 and bZIP17, bZIP28, bZIP49 and bZIP60, respectively), all of which belonged to the virNOG01396 and euNOG19243 groups, respectively (Additional file 1). For phylogenetic analyses, the predicted amino acid sequences were aligned using MUSCLE module [101] with MEGA v.6 software [102]. Unrooted phylogenetic trees were constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA

v.6 software. The trees were visualized with Figtree v1.4 software (http://tree.bio.ed.ac.uk/software/figtree/).

Plasmid construction

The N-terminal cytoplasmic domain of Glyma.02G161100, which spans from bp positions 1 to 627 in the cDNA or from amino acid positions 1 to 209 in the predicted primary structure (Additional file 4C), was isolated from soybean (cv. Conquista) cDNA via PCR using specific primers with appropriate extensions for cloning with Gateway (Life Technologies) (Additional file 15). Similarly, the cytoplasmic domain of glyma03g27865, spanning from bp positions 1 to 1302 or from amino acid positions 1 to 434 (Additional file 4A), and of Glyma.19G126800, spanning from bp positions 1 to 1218 or amino acid positions 1 to 406 (Additional file 4B), were isolated via PCR using gene-specific primers (Additional file 15)

The amplified products were examined by electrophoresis on 1% (w/v) agarose gels, purified using a Gel Extraction Kit (Qiagen) and inserted by recombination into an entry vector, pDonR207 (Life Technologies). The resulting clones pUFV2325 (Glyma.02G161100-pdonR207), pUFV2506 (glyma03g27865-pdonR207) and pUFV2423 (Glyma.19G126800-pdonR207), contained the fragments of the indicated genes covering the N-terminal domain-encoding region up to the transmembrane segment (Additional file 4). These pDonR207-derived clones were used to transfer the respective inserts to a plant expression vector, pEarleyGate-104, generating pUFV2554 (Glyma.02G161100-pEarleyGate-104), pUFV2555 (glyma03g27865-pEarleyGate-104) and pUFV2556 (Glyma.19G126800-pEarleyGate-104), which contained the respective truncated cDNA fragment fused to the C-terminus of yellow fluorescent protein (YFP***)*** under the control of the 35S promoter.

Plant Materials

The tobacco *(N. tabacum* cv. Havana*)* transgenic line BIP-9::GUS has been previously described [42]. The BIP-9::GUS line harbors the promoter (2000 bp of the 5'-flanking region) of the genomic BIP-9 clone (soyBIP cDNA) fused to the reporter gene GUS contained in a plant binary expression vector, pCAMBIA1381z. The vector pCAMBIA1381z contains the *hptII* gene, which confers resistance to

hygromycin. The pCambia::GUS transgenic line, harboring an empty pCAMBIA1381z vector with a promoterless GUS gene, was used as a negative control [42]. Seeds from BiP9::GUS and pCambia::GUS lines were germinated *in vitro* in MS [103] medium supplemented with 25 $mg.L^{-1}$ hygromycin, and they were maintained in a growth chamber at 22°C under a 16 h light/8 h dark cycle for 16 days. Then, the seedlings were transferred to 40-mL pots containing MS medium and hygromycin. At 35 days post-germination, the plants were transferred to the commercial substrate Tropstrato HT and were maintained in growth chambers at 22°C under a 16-h photoperiod for 42 days for transactivation GUS assays.

Transient expression in Nicotiana benthamiana leaves and Nicotiana tabacum cv. Havana transgenic lines by agroinoculation

The *Agrobacterium tumefaciens* strain GV3101 carrying Glyma.02G161100-pEarleyGate-104 (pUFV2554), glyma03g27865-pEarleyGate-104 (pUFV2555) or Glyma.19G126800-pEarleyGate-104 (pUFV2556) DNA constructs was grown for 12 h and subsequently centrifuged for 5 min at 5,000 x *g*. Pelleted cells were washed with 1 mL of infiltration medium (10 mM $MgCl_2$, 10 mM MES, pH 5.6, and 100 µM acetosyringone) and concentrated to an OD_{600nm} = 0.5. *Agrobacterium* infiltration was performed with 3-week-old *N. benthamiana* leaves and 42-day-old BIP-9::GUS and pCambia::GUS transgenic leaves using sterile syringes under manually controlled pressure, as previously described (Batoko et al., 2000 Latijnhouwers et al., 2005). After 72 h, infiltrated leaves from *N. benthamiana* were examined by confocal microscopy, and those from *N. tabacum* cv. Havana transgenic lines were used for GUS transactivation assays. *Agrobacterium* transformed with an unrelated DNA construct, At2g41020 – AtWWP1 (pUFV2224), was used as a control for nuclear localization.

Subcellular localization assay

To examine the subcellular localization of proteins, *N. benthamiana* leaves were agroinoculated with Glyma.02G161100-pEarleyGate-104 (pUFV2554), glyma03g27865-pEarleyGate-104 (pUFV2555) or Glyma.19G126800-pEarleyGate-104 (pUFV2556). These DNA constructs were also co-infiltrated with the nuclear marker *Arabidopsis thaliana* AtWWP1 fused to mCherry (pUFV2224). At

approximately 72 h post-agroinfiltration, 1 cm^2 leaf explants were excised, and YFP and mCherry fluorescence patterns were examined in epidermal cells with a 40x oil immersion objective and a Zeiss LSM510 META inverted laser scanning microscope equipped with argon/helium-neon lasers as excitation sources. For multi-track imaging, YFP was excited with a 488-nm wavelength, and the emission was collected using a 500–530 nm band-pass filter, and mCherry was excited with a 543 nm wavelength, and the emission was collected using a 596-638 nm band-pass filter. The pinhole was typically set to create a 1-1.5-μm optical slice. Post-acquisition image processing was performed using LSM Image Browser 4 software (Carl-Zeiss) and Adobe Photoshop (Adobe Systems).

GUS activity assays

Leaves from BIP-9::GUS and pCambia::GUS transgenic lines were agroinoculated with Glyma.02G161100-pEarleyGate-104 (pUFV2554), glyma03g27865-pEarleyGate-104 (pUFV2555) or Glyma.19G126800-pEarleyGate-104 (pUFV2556) to express the truncated bZIP proteins. Infiltration with *Agrobacterium tumefaciens* carrying a GFP protein (pUFV1088) expression cassette was used as a control. Untransformed, wild-type leaves were also used as a negative control. At 72 h post-infiltration of the *Agrobacterium* suspension cultures, total leaf protein was extracted, and fluorometric assays of GUS activity were performed essentially as described by Jefferson et al. [104], with methylumbelliferone (MU) as a standard. For the standard assay, leaf tissue was ground in 0.5 mL GUS assay buffer (100 mM $NaH_2PO_4 \cdot H_2O$ [pH 7.0], 10 mM EDTA, 0.1% [w/v] sarkosyl, and 0.1% [v/v] Triton X-100), and 50 μL of this extract were mixed with 50 μL GUS assay buffer containing 2 mM fluorescent 4-methylumbelliferyl β-D glucuronide (MUG) as a substrate. The mixture was incubated at 37°C in the dark for 30 min, and GUS activity was measured using a Lector VICTOR™ X5 Multilabel Plate Reader (Perkin Elmer). The total protein concentration was determined by the Bradford method [105]. The experiments were repeated three times with similar results.

Induction of salt stress and ER stress

For the stress treatments, soybean seeds (cv. Conquista) were germinated in an organic substrate and grown under greenhouse conditions. Fifteen days after

germination (V2 stage), roots were washed with water, and the plants were transferred to 200 mmol/L NaCl for 1, 2, 4 and 8 h or 2.5 μg/mL tunicamycin for 4 h for ER stress induction. After all of the treatments, the plant materials were harvested, immediately frozen in liquid N_2 and stored at -80 °C until use. Each stress treatment and RNA extraction were replicated in three independent samples.

Isolation of total RNA from soybean leaves and synthesis of cDNA

Total RNA was extracted from frozen leaves with TRIzol (Invitrogen), according to the manufacturer's instructions. RNA quality and integrity were monitored by electrophoresis on denaturing 1.2% (w/v) agarose gels stained with 0.1 μg/mL ethidium bromide. First-strand cDNA was synthesized from 3 μg RNase-free DNase I-treated total RNA using oligo-dT primers (18) and M-MLV Reverse Transcriptase (Life Technologies), according to the manufacturer's instructions.

Quantitative RT-PCR

The real-time PCR procedures, including the pilot tests, validations, and experiments, were performed according to the information supplied by the Life Technologies manual. Real-time RT-PCR assays were performed with an ABI 7500 instrument (Life Technologies) using SYBR Green PCR Master Mix (Life Technologies) and gene-specific primers (Additional file 16). The conditions for the amplification reactions were as follows: 10 min at 95 °C, followed by 40 cycles at 94 °C for 15 s and 60 °C for 1 min. Absolute gene expression was quantified using the comparative Ct (2–ΔCt) method. Expression of soybean genes was normalized to that of an RNA helicase endogenous control gene, and expression of N. tabacum genes was normalized to that of actin.

Analysis of stress-induced splicing of GmbZIP60 mRNA.

The presence of hairpin-like structures as potential IRE substrates in the sequences of soybean bZIP60 homologs was examined using RNA folding software in Mfold web server version 2.3 (http://mfold.rna.albany.edu/) [106], with the default parameters. This *in silico* analysis revealed a possible functional spliced bZIP60 transcript derived from the Glyma02g19754 sequence using two sets of primers

(Additional files 7 and 15) that were capable of discriminating between spliced and unspliced bZIP60 transcripts in RT-PCR assays. The set of primers glyma02g19754Fwd and glyma02g19754spdR was used for identification of the spliced bZIP60 transcript, whereas the set of primers glyma02g19754Fwd and glyma02g19754unsp amplified the unspliced bZIP60 transcript from cDNA prepared from tunicamycin-treated and untreated total leaf RNA, respectively. Approximately 1 µl of cDNA was used to PCR-amplify the spliced/unspliced transcripts using the indicated pair of primers at a concentration of 0.4 µM each, with 0.2 mM dNTPs, 5 µl of 10X High Fidelity Platinum Taq buffer (Life Technologies) and 0.2 U of High Fidelity Platinum Taq (Life Technologies) in a final volume of 50 µl. The amplification reaction was conducted with an initial denaturation step at 94°C for 3 min, followed by 28 cycles at 94°C for 45 s, 55°C for 45 s, and 68°C for 45 s and a final extension at 68°C for 10 min. The amplification products were examined with a silver-stained 15% (w/v) polyacrylamide gel.

Statistical analyses

All statistical analyses were performed using R software (http://cran.r-project.org) with the ExpDest package [107]. Statistical analysis of GUS activity data was performed using two-way ANOVA (2x6 factorial design, with a completely randomized design and three repetitions) followed by the Scott-Knott test at a $p<0.05$. For the qRT-PCR data, the means were compared using confidence intervals generated by the t test at a $p\leq0.05$.

List of abbreviations

ER, endoplasmic reticulum UPR, unfolded protein response ERAD, endoplasmic reticulum protein degradation IRE1, inositol-requiring transmembrane kinase and endonuclease 1α BiP, binding protein PDI, protein disulfide isomerase CNX, calnexin CRT, calreticulin TF, transcription factor PCD, programmed cell death NRP/DCD, developmental cell death (DCD) domain-containing N-rich protein (NRP) GmERD15, *Glycine max* early responsive to dehydration 15 VPE, vacuolar processing enzyme NAC, no apical meristem (NAM), Arabidopsis ATAF1/2, and cup-shaped cotyledon (CUC).

Competing interests

The authors declare that they have no competing interests

Authors' contributions

PAS carried out the molecular genetics and functional studies, participated in the *in silico* analysis and drafted the manuscript. JCFS and MDBLC carried out the in silico analysis. HDNC participated in the molecular genetics and functional studies of PUR transducers. JPBM participated in the design of the study and performed the confocal microscopy. GCM and PABR participated in the design of the study related to the cell death response. OJBB conducted the statistical analyses of the data. EPBF conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the following grants provided by Brazilian governmental agencies: CNPq grants 573600/2008-2 and 447578/2014-6 (to E.P.B.F.) and an FAPEMIG grant CBB-APQ-00070-09 (to E.P.B.F.). P.A.S., J.C.F.S. and J.P.B.M. were supported by graduate fellowships from CNPq. G.C.M and P.A.B.R. were supported by pos-doctoral fellowships from Fapemig and CNPq, respectively and H.D.N.C was the recipient of a science initiation scholarship from CNPq. We are grateful to the Microscopy Analysis Core of the Universidade Federal de Vicosa for providing the facilities for confocal microscopy. We also thank Professors Murilo Zerbini and Anésia A. Santos from the Universidade Federal de Viçosa and Professor Carlos Priminho Pirovani from the Univesidade Estadual Santa Cruz for the critical reading of the manuscript.

REFERENCES

1. Ron D, Walter P: **Signal integration in the endoplasmic reticulum unfolded protein response.** *Nat Rev Mol Cell Biol* 2007, **8:**519–529.
2. Schröder M: **Endoplasmic reticulum stress responses.** *Cell Mol Life Sci* 2008, **65**:862-894.
3. Sparkes IA, Frigerio L, Tolley N, Hawes C: **The plant endoplasmic reticulum: a cell-wide web.** *Biochem J* 2009, **423**:145–155.
4. Malhotra JD, Kaufman RJ: **The endoplasmic reticulum and the unfolded protein response.** *Semin Cell Dev Biol* 2007, **18:**716–731.
5. Määttänen P, Gehring K, Bergeron JJM, Thomas DY: **Protein quality control in the ER: The recognition of misfolded proteins.** *Semin Cell Dev Biol* 2010, **21:**500–511.
6. Walter P, Ron D: **The unfolded protein pesponse: from stress pathway to homeostatic regulation.** *Science.* 2011, **334:**1081–1086.
7. Liu J, Howell SH: **Endoplasmic reticulum protein quality control and its relationship to environmental stress responses in plants.** *Plant Cell* 2010, **22:**2930–2942.
8. Deng Y, Srivastava R, Howell SH: **Endoplasmic reticulum (RE) stress response and its physiological roles in plants**. *Int J Mol Sci* 2013, **14:**8188–8212.
9. Howell SH: **Endoplasmic reticulum stress responses in plants**. *Annu Rev Plant Biol* 2013, **64:**477–499.
10. Yang ZT, Lu SJ, Wang MJ, Bi DL, Sun L, Zhou SF, Song ZT, Liu JX: **A plasma membrane-tethered transcription factor, NAC062/ANAC062/NTL6, mediates the unfolded protein response in Arabidopsis.** *Plant J* 2014, **79:**1033–1043.
11. Hayashi S, Wakasa Y, Takahashi H, Kawakatsu T, Takaiwa F: **Signal transduction by IRE1-mediated splicing of bZIP50 and other stress sensors in the endoplasmic reticulum stress response of rice**. *Plant J* 2012, **69**:946–956.
12. Koizumi N, Martinez, IM, Kimata Y, Kohno K, Sano H, Chrispeels MJ: **Molecular characterization of two arabidopsis Ire1 transmembrane protein kinases.** *Plant Physiol* 2001, **127:**949–962.

13. Li Y, Humbert S, Howell SH: **ZmbZIP60 mRNA is spliced in maize in response to ER stress.** *BMC Res. Notes* 2012, **5**:144.
14. Deng Y, Humbert S, Liu J-X, Srivastava R, Rothstein SJ, Howell SH: **Heat induces the splicing by IRE1 of a mRNA encoding a transcription factor involved in the unfolded protein response in Arabidopsis.** *Proc Natl Acad Sci USA* 2011, **108:**7247–7252.
15. Nagashima Y, Mishiba K, Suzuki E, Shimada Y, Iwata Y, Koizumi N: **Arabidopsis IRE1 catalyses unconventional splicing of bZIP60 mRNA to produce the active transcription factor.** *Sci. Rep.* 2011, **1**:29.
16. Oikawa D, Tokuda M, Hosoda A, Iwawaki T: **Identification of a consensus lement recognized and cleaved by IRE1a.** *Nucleic Acids Res* 2010, **38:**6265–6273.
17. Iwata Y, Koizumi N: **Plant transducers of the endoplasmic reticulum unfolded protein response.** *Trends Plant Sci* 2012, **17:**720–727.
18. Liu L, Cui F, Li Q, Yin B, Zhang H, Lin B, Wu Y, Xia R, Tang S, Xie Q: **The endoplasmic reticulum-associated degradation is necessary for plant salt tolerance.** *Cell Res* 2011, **21:**957–969.
19. Liu J-X, Srivastava R, Che P, Howell SH: **An endoplasmic reticulum stress response in Arabidopsis is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28.** *Plant Cell* 2007, **19:**4111–4119.
20. Srivastava R, Deng Y, Shah S, Rao AG, Howell SH: **BINDING PROTEIN is a master regulator of the endoplasmic reticulum stress sensor/transducer bZIP28 in Arabidopsis.** *Plant Cell* 2013, **25:**1416–1429.
21. Sun L, Zhang SS, Lu SJ, Liu JX: **Site-1 protease cleavage site is important for the ER stress-induced activation of membrane-associated transcription factor bZIP28 in Arabidopsis**. *Sci China Life Sci* 2015, 58:270-275.
22. Srivastava R, Deng Y, Howell SH: **Stress sensing in plants by an ER stress sensor/transducer, bZIP28.** *Front Plant Sci* 2014, **5:** 59.
23. Liu JX, Howell SH: **bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis.** *Plant Cell* 2010, **22:**782–796.
24. Henriquez-Valencia C, Moreno AA, Sandoval-Ibañez O, Mitina I, Blanco-Herrera F, Cifuentes-Esquivel N, Orellana A: **bZIP17 and bZIP60 regulate the**

expression of BiP3 and other salt stress responsive genes in an UPR-independent manner in Arabidopsis thaliana. *J Cell Biochem* 2015, doi: 10.1002/jcb.25121

25. Liu JX, Srivastava R, Howell SH: **Stress-induced expression of an activated form of AtbZIP17 provides protection from salt stress in Arabidopsis**. *Plant Cell Environ* 2008, **31:**1735–1743.
26. Liu JX, Srivastava R, Che P, Howell SH: **Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling.** *Plant J* 2007, **51:**897–909.
27. Gao H, Brandizzi F, Benning C, Larkin RM: **A membrane-tethered transcription factor defines a branch of the heat stress response in Arabidopsis thaliana.** *Proc Natl Acad Sci U S A* 2008, **105:**16398–16403.
28. Yang YG, Lv WT, Li MJ, Wang B, Sun DM, Deng X: **Maize membrane-bound transcription factor zmbzip17 is a key regulator in the cross-talk of er quality control and aba signaling**. *Plant Cell Physiol* 2013, **54**:2020–2033.
29. Che P, Bussell JD, Zhou W, Estavillo GM, Pogson BJ, Smith SM: **Signaling from the endoplasmic reticulum activates brassinosteroid signaling and promotes acclimation to stress in Arabidopsis.** *Sci. Signal.* 2010, **3:**ra69.
30. Fanata WID, Lee SY, Lee KO: **The unfolded protein response in plants: A fundamental adaptive cellular response to internal and external stresses**. *J Proteomics* 2013, **93**:356-368.
31. Sun L, Yang ZT, Song ZT, Wang MJ, Sun L, Lu SJ, Liu JX: **The plant-specific transcription factor gene NAC103 is induced by bZIP60 through a new cis-regulatory element to modulate the unfolded protein response in Arabidopsis.** *Plant J* 2013, **76:**274–286.
32. Kamauchi S, Nakatani H, Nakano C, Urade R: **Gene expression in response to endoplasmic reticulum stress in *Arabidopsis thaliana***. *FEBS J.* 2005, **272**:3461–3476.
33. Kirst ME, Meyer DJ, Gibbon BC, Jung R, Boston RS: **Identification and characterization of endoplasmic reticulum-associated degradation proteins differentially affected by endoplasmic reticulum stress.** *Plant Physiol* 2005, **138:**218–231.
34. Liu Y, Li J: **Endoplasmic reticulum-mediated protein quality control in Arabidopsis.** *Front. Plant Sci* 2014, **5:**162.

35. Martinez IM, Chrispeels MJ: **Genomic Analysis of the unfolded protein response in Arabidopsis shows its connection to important cellular processes**. *Plant Cell* 2003, **15:**561–576.
36. Yang ZT, Wang MJ, Sun L, Lu SJ, Bi DL, Sun L, Song ZT, Zhang SS, Zhou SF, Liu JX: **The membrane-associated transcription factor NAC089 controls ER-stress-induced programmed cell death in plants.** *PLoS Genet* 2014, **10**:e1004243.
37. Costa MDL, Reis PAB, Valente MAS, Irsigler AST, Carvalho CM, Loureiro ME, Aragão FJL, Boston RS, Fietto LG, Fontes EPB: **A new branch of endoplasmic reticulum stress signaling and the osmotic signal converge on plant-specific asparagine-rich proteins to promote cell death**. *J Biol Chem* 2008, **283:**20209–20219.
38. Alves MS, Reis PAB, Dadalto SP, Faria JAQA, Fontes EPB, Fietto LG: **A novel transcription factor, ERD15 (Early Responsive to Dehydration 15), connects endoplasmic reticulum stress with an osmotic stress-induced cell death signal.** *J Biol Chem* 2011, **286**:20020–20030.
39. Faria JAQA, Reis PAB, Reis MTB, Rosado GL, Pinheiro GL, Mendes GC, Fontes EPB: **The NAC domain-containing protein , GmNAC6 , is a downstream component of the ER stress- and osmotic stress-induced NRP-mediated cell-death signaling pathway.** *BMC Plant Biol* 2011, **11:**129.
40. Mendes GC, Reis PAB, Calil IP, Carvalho HH, Aragão FJL, Fontes EPB: **GmNAC30 and GmNAC81 integrate the endoplasmic reticulum stress- and osmotic stress-induced cell death responses through a vacuolar processing enzyme.** *Proc Natl Acad Sci USA* 2013, **110:**19627-19632.
41. Irsigler AST, Costa MDL, Zhang P, Reis PAB, Dewey RE, Boston RS, Fontes EPB: **Expression profiling on soybean leaves reveals integration of ER- and osmotic-stress pathways.** *BMC Genomics* 2007, **8:**431.
42. Buzeli AAR, Cascardo JCM, Rodrigues LAZ, Andrade MO, Almeida RS, Loureiro ME, Otoni WC, Fontes EPB: **Tissue-specific regulation of *BiP* genes: a cis-acting regulatory domain is required for *BiP* promoter activity in plant meristems.** *Plant Mol Biol* 2002, **50**:757–771.
43. Carvalho HH, Silva PA, Mendes GC, Brustolini OJB, Pimenta MR, Gouveia BC, Valente MAS, Ramos HJO, Ramos JRLS, Fontes EPB: **The endoplasmic**

reticulum binding protein bip displays dual function in modulating cell death events. *Plant Physiol* 2014, **164:**654–670.

44. Schmutz J et al: **Genome sequence of the palaeopolyploid soybean**. *Nature* 2010, **463:**178–183.
45. Srivastava R, Chen Y, Deng Y, Brandizzi F, Howell SH: **Elements proximal to and within the transmembrane domain mediate the organelle-to-organelle movement of bZIP28 under ER stress conditions.** *Plant J* 2012, **70:**1033–1042.
46. De Craene JO, Courte F, Rinaldi B, Fitterer C, Herranz MC, Schmitt-Keichinger C, Ritzenthaler C, Friant S: **Study of the plant COPII vesicle coat subunits by functional complementation of yeast Saccharomyces cerevisiae mutants**. *PLoS One* 2014, **9:**e90072.
47. Hanton SL, Chatre L, Matheson LA, Rossi M, Held MA, Brandizzi F: **Plant Sar1 isoforms with near-identical protein sequences exhibit different localisations and effects on secretion.** *Plant Mol Biol* 2008, **67:**283-294.
48. Carvalho HH, Brustolini OJ, Pimenta MR, Mendes GC, Gouveia BC, Silva PA, Silva JC, Mota CS, Soares-Ramos JR, Fontes EPB: **The molecular chaperone binding protein BiP prevents leaf dehydration-induced cellular homeostasis disruption**. *PLoS One* 2014, **9:**e86661.
49. Iwata Y, Koizumi N: **An Arabidopsis transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants**. *Proc Natl Acad Sci USA* 2005, **12:**5280-5285.
50. Iwata Y, Fedoroff NV, Koizumi N: **Arabidopsis bZIP60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response.** *Plant Cell* 2008, **20:**3107–3121.
51. Noh S, Kwon CS, Chung WI: **Characterization of two homologs of Ire1p, a kinase/endoribonuclease in yeast, in *Arabidopsis thaliana*.** *Biochim. Biophys Acta* 2002, **1575:**130–134.
52. Oh D, Kwon C, Sano H, Chung W, Koizumi N: **Conservation between animals and plants of the cis-acting element involved in the unfolded protein response**. *Biochem Biophys Res Commun* 2003, **30:**225–230.
53. Hayashi S, Takahashi H, Wakasa Y, Kawakatsu T, Takaiwa F: **Identification of a cis-element that mediates multiple pathways of the endoplasmic reticulum stress response in rice.** *Plant J* 2013, **74:**248–257.

54. Pinheiro GL, Marques CS, Costa MDBL, Reis PAB, Alves MS, Carvalho CM, Fietto LG, Fontes EPB: **Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response**. *Gene* 2009, **444:**10–23.
55. Cascardo JCM, Almeida RS, Buzeli RAA, Carolino SMB, Otoni WC, Fontes EBP: **The phosphorylation state and expression of soybean bip isoforms are ifferentially regulated following abiotic stresses.** *J Biol Chem* 2000, **275:**14494-14500.
56. Cascardo JCM, Buzeli RAA, Almeida RS, Otoni WC, Fontes EPB: **Differential expression of the soybean BiP gene family**. *Plant Sci* 2001, **160:** 273–281.
57. Denecke J, Goldman MHS, Demolder J, Seurinck J, Bottermana J: **The tobacco luminal binding protein is encoded by a multigene family.** *Plant Cell* 1991, **3:**1025–1035.
58. Fontes EBP, Shank BB, Wrobel RL, Moose SP, Obrian GR, Wurtzel ET, Boston RS: **Characterization of an lmmunoglobulin binding protein homolog in the maize floury-2 endosperm mutant.** *Plant Cell* 1991, **3:**483–496.
59. Kalinski A, Rowley DL, Loer DS, Foley C, Buta G, Herman EM: **Binding-protein expression is subject to temporal, developmental and stress-induced regulation in terminally differentiated soybean organs**. *Planta* 1995, **195**:611–621.
60. Lin B, Wang J, Liu H, Chen R, Meyer Y, Barakat A, Delseny M: **Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana.** *Cell Stress Chaperones* 2001, **6:**201–208.
61. Alvim FC, Carolino SMB., Cascardo JCM, Nunes CC, Martinez CA, Otoni WC, Fontes EPB: *Enhanced accumulation of bip in transgenic plants confers tolerance to water stress. Plant Physiol* 2001, **126:**1042–1054.
62. Leborgne-Castel N, Jelitto-Van Dooren EPWM, Crofts AJ, Denecke J: **Overexpression of BiP in tobacco alleviates endoplasmic reticulum stress.** *Plant Cell* 1999, **11:**459–470.
63. Valente MAS, Faria JAQA, Ramos JRLS, Reis PAB, Pinheiro GL, Piovesan ND, Morais T, Menezes CC, Cano MAO, Fietto LG, Loureiro ME, Aragão FJL, Fontes EPB: **The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco.** *J Exp Bot* 2009, **60:**533–546.

64. Wang D, Weaver ND, Kesarwani M, Dong X: **Induction of protein secretory pathway is required for systemic acquired resistance.** *Science* 2005, **308**:1036–1040.
65. Reis PAB, Rosado GL, Silva LAC, Oliveira LC, Oliveira LB, Costa MDBL, Alvim FC, Fontes EPB: **The binding protein BiP attenuates stress-induced cell death in soybean via modulation of the N-Rich protein-mediated signaling pathway.** *Plant Physiol* 2011, **157:**1853–1865.
66. Jin H, Yan Z, Nam KH, Li J: **Allele-specific suppression of a defective brassinosteroid receptor reveals a physiological role of UGGT in ER quality control**. *Mol Cell* 2007, **26**:821-830.
67. Jia X, He L, Jing R, Li R: **Calreticulin : conserved protein and diverse functions in plants.** *Physiol Plant* 2009, **136:**127–138.
68. Aller I, Meyer AJ: **The oxidative protein folding machinery in plant cells.** *Protoplasma* 2013, **250:** 799–816.
69. Houston NL, Fan C, Xiang JQ, Schulze JM, Jung R, Boston RS: **Phylogenetic analyses identify 10 classes of the protein disulfide isomerase family in plants, including single-domain protein disulfide isomerase-related proteins.** *Plant Physiol* 2005, **137:**762–778.
70. Selles B, Jacquot JP, Rouhier N: **Comparative genomic study of protein disulfide isomerases from photosynthetic organisms**. *Genomics* 2011, **97:**37–50.
71. Landrieu I, De Veylder L, Fruchart JS, Odaert B, Casteels P, Portetelle D, Van Montagu M, Inzé D, Lippens G: **The *Arabidopsis thaliana* PIN1At gene encodes a single-domain phosphorylation-dependent peptidyl prolyl cis/trans isomerase.** *J Biol Chem* 2000, **275:**10577–10581.
72. Yao JL, Kops O, Lu PJ, Lu KP: **Functional conservation of phosphorylation-specific prolyl isomerases in plants**. *J Biol Chem* 2001, **276:**13517–13523.
73. Schlenstedt G, Harris S, Risse B, Lill R, Silver PA: **A yeast DnaJ homologue, Scjlp, can function in the endoplasmic reticulum with BiP/Kar2p via a conserved domain that specifies interactions with Hsp70s**. *J Cell Biol* 1995, **129:**979–988.
74. Klein EM, Mascheroni L, Pompa A, Ragni L, Weimar T, Lilley KS, Dupree P, Vitale A: **Plant endoplasmin supports the protein secretory pathway and has a role in proliferating tissues**. *Plant J* 2006, **48:**657–673.

75. Koiwa H, Li F, McCully MG, Mendoza I, Koizumi N, Manabe Y, Nakagawa Y, Zhu J, Rus A, Pardo JM, Bressan RA, Hasegawa PM: **The STT3a subunit isoform of the Arabidopsis oligosaccharyltransferase controls adaptive responses to salt/osmotic stress.** *Plant Cell* 2003, **15:**2273–2284.
76. Matsushima R, Kondo M, Nishimura M, Hara-Nishimura I: **A novel ER-derived compartment, the ER body, selectively accumulates a beta-glucosidase with an ER-retention signal in Arabidopsis.** *Plant J* 2003, **33:**493-502.
77. Burn JE, Hurley UA, Birch RJ, Arioli T, Cork A, Williamson RE: **The cellulose-deficient Arabidopsis mutant rsw3 is defective in a gene encoding a putative glucosidase II, an enzyme processing N-glycans during ER quality control.** *Plant J* 2002, **32:**949–960.
78. Jin H, Hong Z, Su W, Li J: **A plant-specific calreticulin is a key retention factor for a defective brassinosteroid receptor in the endoplasmic reticulum**. *Proc Natl Acad Sci USA* 2009, **106:**13612–13617.
79. Lemus L, Goder V: **Regulation of endoplasmic reticulum-associated protein degradation (ERAD) by ubiquitin.** *Cells* 2014, **3:**824–847.
80. Hüttner S, Veit C, Schoberer J, Grass J, Strasser R: **Unraveling the function of Arabidopsis thaliana OS9 in the endoplasmic reticulum-associated degradation of glycoproteins.** *Plant Mol Biol* 2012, **79:**21–33.
81. Hong Z, Jin H, Fitchette A, Xia Y, Monk AM, Li J: **Mutations of an a 1,6 mannosyltransferase inhibit endoplasmic reticulum–associated degradation of defective brassinosteroid receptors in arabidopsis.** *Plant Cell* 2009, **21:**3792–3802.
82. Hong Z, Jin H, Tzfira T, Li J: **Multiple mechanism-mediated retention of a defective brassinosteroid receptor in the endoplasmic reticulum of Arabidopsis.** *Plant Cell* 2008, **20:**3418–3429.
83. Hong Z, Kajiura H, Su W, Jin H, Kimura A, Fujiyama K, Li J: **Evolutionarily conserved glycan signal to degrade aberrant brassinosteroid receptors in Arabidopsis**. *Proc Natl Acad Sci USA* 2012, **109:**11437-11442.
84. Liebminger E, Hüttner S, Vavra U, Fischl R, Schoberer J, Grass J, Blaukopf, C, Seifert GJ, Altmann F, Mach L, Strasser R: **Class I alpha-mannosidases are required for N-glycan processing and root development in *Arabidopsis thaliana.*** *Plant Cell* 2009, **21:**3850–3867.

85. Su W, Liu Y, Xia Y, Hong Z, Li J: **The arabidopsis homolog of the mammalian os-9 protein plays a key role in the endoplasmic reticulum-associated degradation of misfolded receptor-like kinases**. *Mol Plant* 2012, **5:**929–940.
86. Su W, Liu Y, Xia Y, Hong Z, Li J: **Conserved endoplasmic reticulum-associated degradation system to eliminate mutated receptor-like kinases in Arabidopsis.** *Proc Natl Acad Sci USA* 2011, **108:**870–875.
87. Doblas VG, Amorim-Silva V, Posé D, Rosado A, Esteban A, Arró M, Azevedo H, Bombarely A, Borsani O, Valpuesta V, Ferrer A, Tavares RM, Botella MA: **The SUD1 gene encodes a putative E3 ubiquitin ligase and is a positive regulator of 3-hydroxy-3-methylglutaryl coenzyme a reductase activity in Arabidopsis.** *Plant Cell* 2013, **25:**728–743.
88. Cui F, Liu L, Zhao Q, Zhang Z, Li Q, Lin B, Wu Y, Tang S, Xie Q: **Arabidopsis ubiquitin conjugase UBC32 is an ERAD component that functions in brassinosteroid-mediated salt stress tolerance.** *Plant Cell* 2012, **24:**233–244.
89. Kaur N, Zhao Q, Xie Q, Hu J: **Arabidopsis ring peroxins are e3 ubiquitin ligases that interact with two homologous ubiquitin receptor proteins.** *J Integr Plant Bio*l 2013, **55:**108–120.
90. Kraft E, Stone SL, Ma L, Su N, Gao Y, Lau OS, Deng XW, Callis J: **Genome analysis and functional characterization of the E2 and RING-type E3 ligase ubiquitination enzymes of Arabidopsis**. *Plant Physiol* 2005, **139:**1597–1611.
91. Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH: **Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis.** *Plant Physiol* 2008, **148:**1201–1211.
92. Wolf DH, Stolz A: **The Cdc48 machine in endoplasmic reticulum associated protein degradation**. *Biochim Biophys Acta* 2012, **1823:**117–124.
93. Raasi S, Wolf DH: **Ubiquitin receptors and ERAD : A network of pathways to the proteasome**. *Semin Cell Dev Biol* 2007, **18:**780–791.
94. Rancour DM, Dickey CE, Park S, Bednarek SY: **Characterization of AtCDC48. Evidence for multiple membrane fusion mechanisms at the plane of cell division in plants.** *Plant Physiol* 2002, **130:**1241–1253.
95. Rancour DM, Park S, Knight SD, Bednarek SY: **Plant UBX domain-containing protein 1, PUX1, regulates the oligomeric structure and activity of arabidopsis CDC48.** *J Biol Chem* 2004, **279:**54264–54274.

96. Park S, Rancour DM, Bednarek SY: **Protein domain-domain interactions and requirements for the negative regulation of Arabidopsis CDC48/p97 by the plant ubiquitin regulatory X (UBX) domain-containing protein, PUX1.** *J Biol Chem* 2007, **282:**5217–5224.
97. Le DT, Nishiyama RIE, Watanabe YA, Mochida KE, Yamaguchi-Shinozaki, K, Shinozaki K, Tran LSP: **Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress.** *DNA Res* 2011, **18:**263–276.
98. Hoeoflinger MC, Pieslinger AM, Tenhaken R: **Plant physiology and biochemistry investigations on N-rich protein (NRP) of Arabidopsis thaliana under different stress conditions.** *Plant Physiol Biochem* 2011, **49:**293-302.
99. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P: **eggNOG v4.0 : nested orthology inference across 3686 organisms**. *Nucleic Acids Res* 2014, **42**(Database issue):D231-239
100. Möller S, Croning MD, Apweiler R: **Evaluation of methods for the prediction of membrane spanning regions.** *Bioinformatics* 2001, **17:**646-653.
101. Edgar RC, Drive RM, Valley M: **MUSCLE : multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792–1797.
102. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6 : Molecular Evolutionary Genetics Analysis Version 6.** *Mol Biol Evol* 2013, **30:**2725–2729.
103. Murashige T, Skoog F: **A revised medium for rapid growth and bioassays with tobacco tissue cultures.** *Physiol Plant* 1962, **15**:473–497.
104. Jefferson RA, Kavanagh TA, Bevan MW: **Gus fusions: *β*-glucuronidase as a sensitive and versatile gene fusion marker in higher plants.** *EMBO J* 1987, **6:**3901–3907.
105. Bradford M: **Photometric methods for protein determination. Procedures and analysis.** *Anal Biochem* 1976, **72:**248-254.
106. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31:**3406–3415.
107. Ferreira EB, Cavalcanti PP, Nogueira DA: **ExpDes : an R package for ANOVA and experimental designs.** *App Math* 2014, **5:**2952–2958.

108. Suzuki, T., Park, H., Till, E. A, and Lennarz, W.J. 2001. **The PUB domain: a putative protein-protein interaction domain implicated in the ubiquitin-proteasome pathway.** *Biochem. Biophys. Res. Commun.* 287: 1083–1087.

109. Galvão, R.M., Kota, U., Soderblom, E.J., Goshe, M.B., and Boss, W.F. 2008. **Characterization of a new family of protein kinases from Arabidopsis containing phosphoinositide 3/4-kinase and ubiquitin-like domains.** *Biochem. J.* 409: 117–127

110. Diepold, A., Li, G., Lennarz, W.J., Nürnberger, T., and Brunner, F. 2007. **The Arabidopsis AtPNG1 gene encodes a peptide: N-glycanase.** *Plant J.* 52: 94–104.

111. Farmer, L.M., Book, A.J., Lee, K.-H., Lin, Y.-L., Fu, H., and Vierstra, R.D. 2010. **The RAD23 family provides an essential connection between the 26S proteasome and ubiquitylated proteins in Arabidopsis.** *Plant Cel.l* 22: 124–142.

112. Chen, Y. and Brandizzi, F. 2012. **AtIRE1A/AtIRE1B and AGB1 independently control two essential unfolded protein response pathways in Arabidopsis.** *Plant J.* 69: 266–277.

113. Wang, S., Narendra, S., and Fedoroff, N. 2007. **Heterotrimeric G protein signaling in the Arabidopsis unfolded protein response.** *Proc. Natl. Acad. Sci. U. S. A.* 104: 3817–3822.

FIGURE LEGENDS

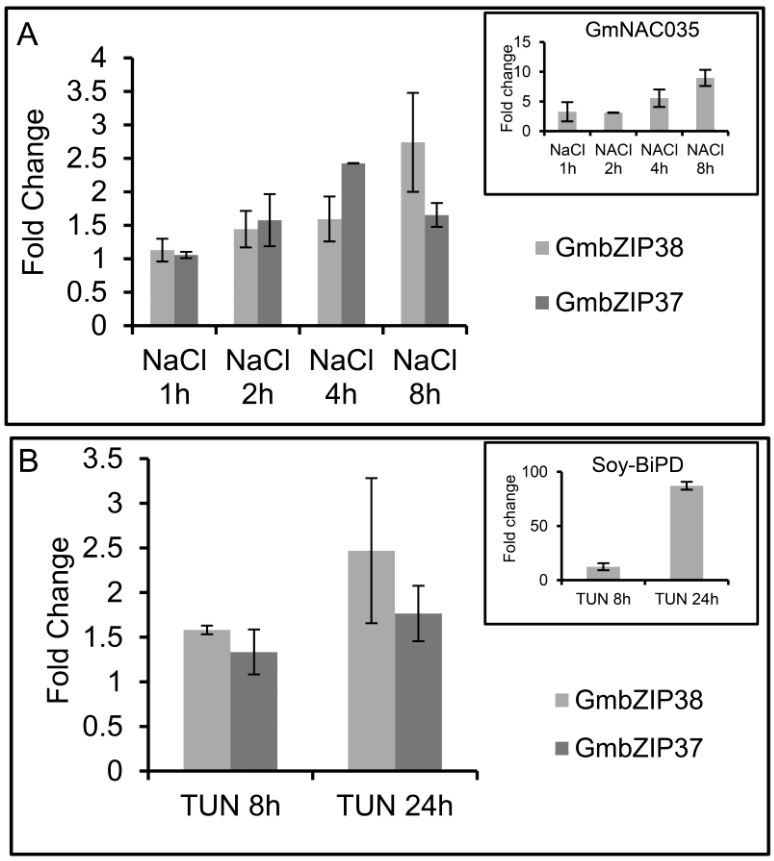


Figure 1. Expression analysis of bZIP38 and bZIP37. Total RNA was isolated from soybean seedlings treated with NaCl (A) and tunicamycin (B) for the indicated times, and the transcript levels of selected genes (as indicated) were quantified by real-time PCR using gene-specific primers. Gene expression was calculated using the $2^{-\Delta CT}$ method, with helicase as an endogenous control. The error bars indicate 95% confidence intervals based on t-tests ($p<0.05$, $n=3$).

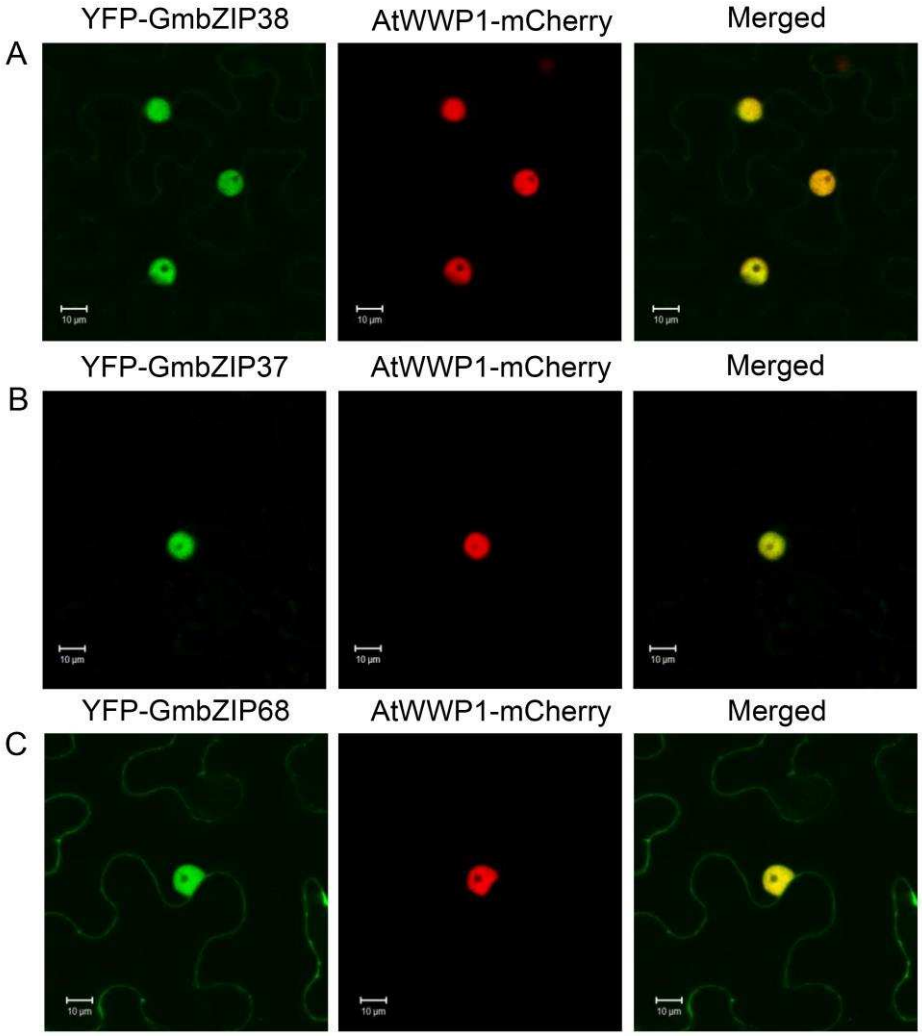


Figure 2. Subcellular localizations of the truncated forms of bZIP38, bZIP37 and bZIP68 fused to YFP. *N. benthamiana* leaves were infiltrated with *Agrobacterium* carrying the indicated DNA constructs. The subcellular localizations of the fluorescent fusion proteins were monitored by confocal microscopy at 72 h post-infiltration. The co-localization of the YFP-bZIP fusion proteins with the nuclear marker AtWWP1-mCherry is shown in the merged image. Scale bars = 10 µm.

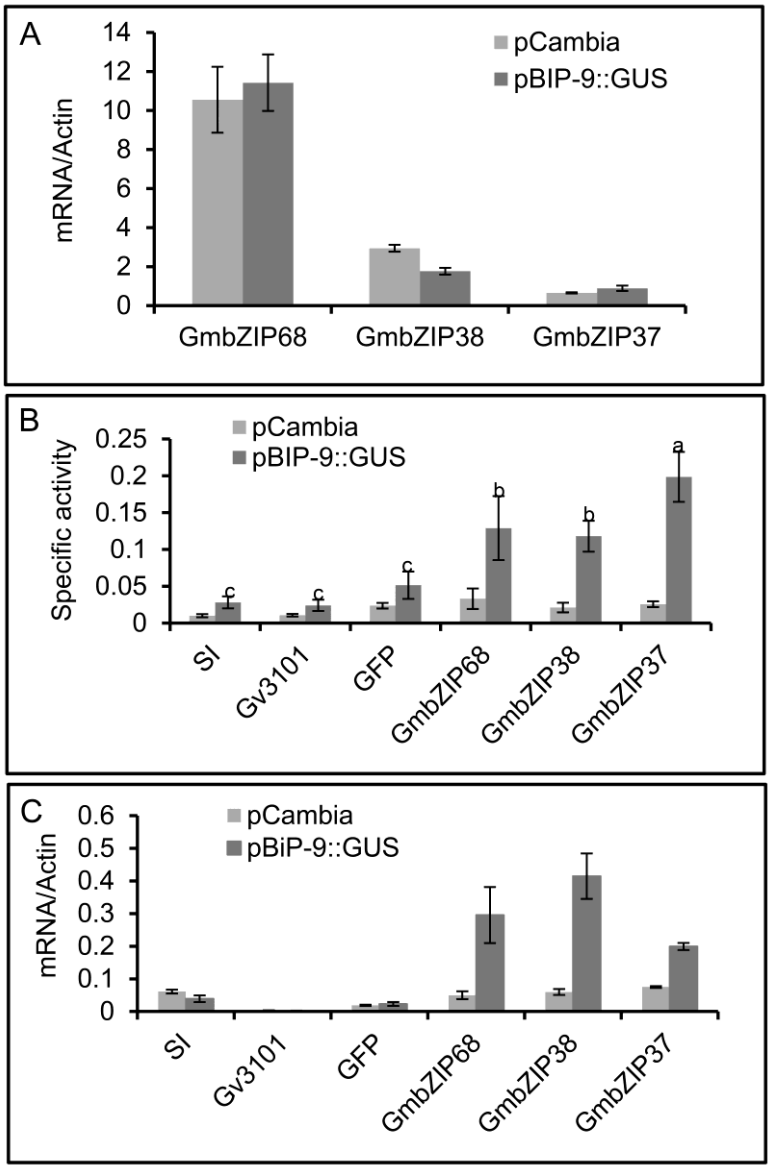


Figure 3. GmbZIP38, GmbZI37 and GmbZIP68 activate a BiP promoter. The leaves of transgenic tobacco lines transformed either with a soyBiPD promoter fused to GUS (pBiP-9::GUS) or an empty pCambia vector (pCambia) were agroinfiltrated with plasmids carrying truncated GmbZIP38 ($bZIP38^{1-434}$), GmbZIP37 ($bZIP37^{1-406}$) and GmbZIP68 ($bZIP68^{1-209}$). A. Expression of truncated bZIPs in agroinfiltrated leaves. The expression levels of truncated GmbZIP38, GmbZIP37 and GmbZIP68 were analyzed by qRT-PCR at 72 h post-infiltration. Expression levels were calculated using the $2^{-\Delta CT}$ method, with helicase as an endogenous control. The error bars indicate 95% confidence intervals based on t-tests ($p<0.05$, n=3). B. Induction of GUS activity in transgenic lines by expression of truncated GmbZIP38, GmbZIP37 and GmbZIP68. Transgenic tobacco leaves (pBiP-9::GUS and pCambia) were infiltrated with *A. tumefaciens* carrying the indicated DNA constructs, and GUS activity was measured at 72 h post-infiltration. Non-inoculated (SI) transgenic lines and those inoculated with either GV3101 or GFP were used as negative controls. C. GUS transcript accumulation. The expression of GUS was analyzed by qRT-PCR at 72 h post-infiltration. Expression levels were calculated using the $2^{-\Delta CT}$ method, and helicase served as an endogenous control. The error bars indicate 95% confidence intervals based on t-tests ($p<0.05$, n=3).

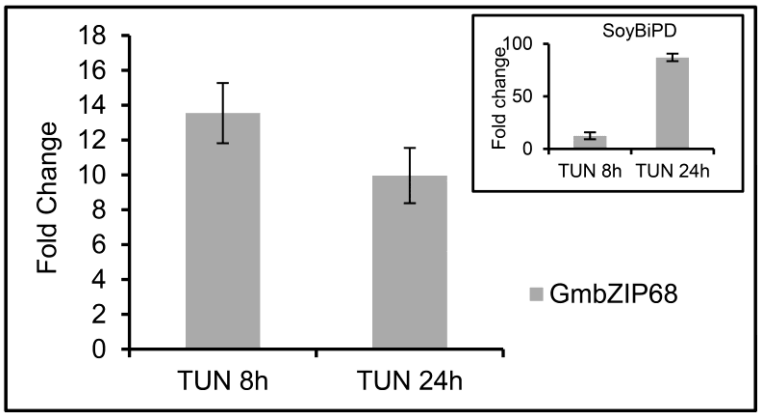


Figure 4. Expression analysis of bZIP68. Total RNA was isolated from soybean seedlings treated with tunicamycin for the indicated time, and the transcript level of the truncated form of bZIP68 was quantified by real-time PCR using gene-specific primers. Gene expression was calculated using the $2^{-\Delta CT}$ method, with helicase as an endogenous control. The error bars indicate 95% confidence intervals based on t-tests ($p<0.05$, $n=3$).

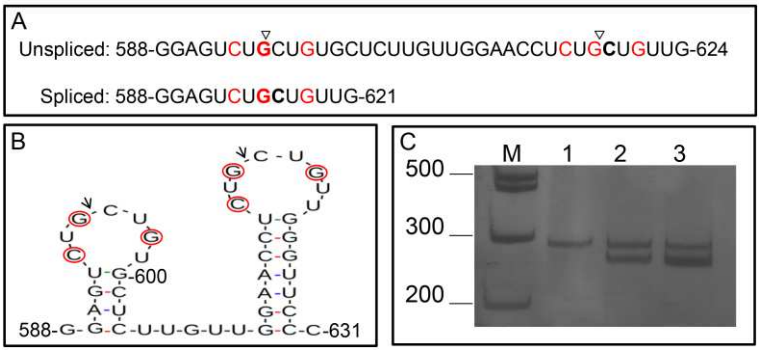


Figure 5. Regulated splicing of bZIP68 caused by ER stress. A. Predicted spliced sequence of the bZIP68 transcript. The arrows indicate the predicted cleavage sites, and the nucleotides in red are conserved in IRE1 substrates. B. Loop structure at the predicted splicing site in bZIP68 mRNA. Each of the two loops contains three conserved nucleotides (marked in red) present in IRE1 substrates. C. ER stress-induced splicing of bZIP68 mRNA. Total RNA was isolated from soybean seedlings that were untreated (lane 1) or treated with tunicamycin for 8 h (lane 2) and 24 h (lane 3) and used as a template for RT-PCR with a combination of primers for spliced and unspliced bZIP68.

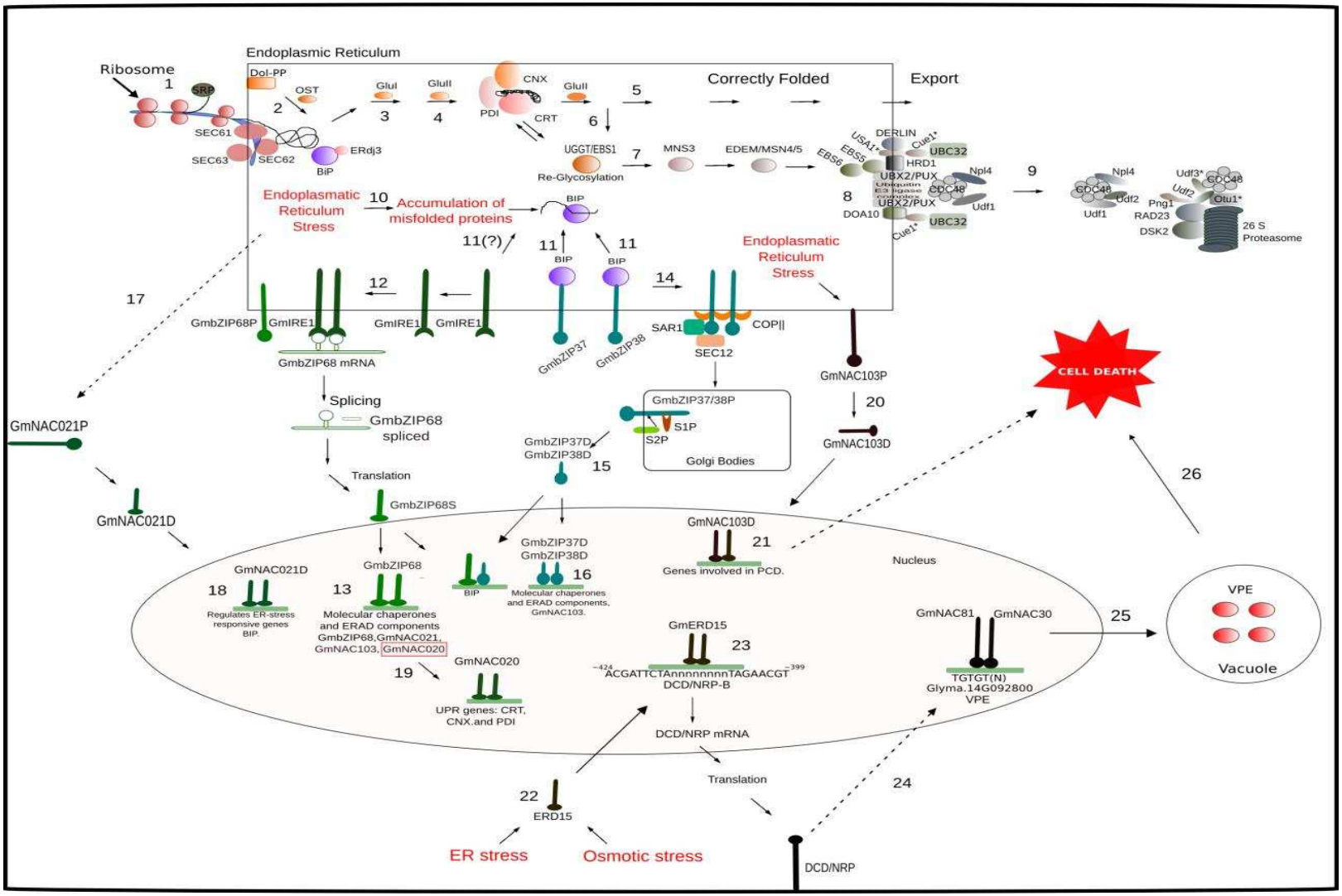


Figure 6. Comprehensive panel of the ER stress response in soybean.

Figure 6. Comprehensive panel of the ER stress response in soybean. To enter the ER lumen, secretory proteins are translated by ER-associated polysomes, and the nascent secretory peptide is co-translationally transported to the ER through the Sec61 translocation complex (1). In the ER, the pre-assembly oligosaccharide core (Glc3Man9GlcNAc2 N-glycan) is transferred (2) from the ER-localized dolichyl pyrophosphate (Dol-PP) to the nascent polypeptide by oligosaccharyltransferase (OST). Processing or trimming of the N-glycan begins in the ER with the sequential removal of the more external glucose residues by glucosidase I (3) and glucosidase II (4). The monoglycosylated glucan-peptide is targeted to the calnexin/calreticulin system containing the protein disulfide isomerase (PDI) accessory protein for proper folding (5). Folded proteins are released from this N-glycan-dependent quality control mechanism through hydrolysis of the third glucose residue by GluII. Properly folded proteins leave the ER. Unfolded proteins may be re-glycosylated by UDP-glucose:glycoprotein glucosyltransferase (UGGT) to re-enter the CNX/CRT-mediated folding cycle (6). The removal of glucose residues and transient re-addition of the innermost glucose during protein folding contribute to the ER retention time of a given glycoprotein. Failure to achieve the proper conformation within a defined period of time is a signal for exclusion of the glycoprotein from the CNX/CRT folding cycle by the sequential removal of two α1,2-mannose residues by MNS3 and MNS4/MNS5 (7). The removal of these residues exposes an α1,6-mannose, which targets the glycoprotein to the ERAD pathway. EBS6 and EBS5 recruit unfolded glycoproteins to redirect them to the membrane-associated ERAD complex for ubiquitination and retrotranslocation to the cytosol, where they are targeted to the proteasome (9). Ubiquitinated ERAD substrates are directed from the ER to the proteasome via the trimeric complex cdc48/Ufd1/Npl4 (9). ER stress induces the accumulation of unfolded proteins in the lumen and activates the UPR pathway (10). BiP-mediated dissociation 0of the UPR transducers GmbZIP37/38 (AtbZIP17/28) (11) allows for the mobilization of these receptors to the Golgi (14), where they are proteolytically cleaved by S1P and S2P (15), releasing the N-terminal bZIP domains as functional TFs that are then translocated to the nucleus (15), where they activate ER stress-responsive promoters (16). In the other arm of the UPR (12), under ER stress, GmIRE1 dimerizes to activate its ribonuclease activity, which promotes unconventional splicing of the GmbZIP68 (AtbZIP60) mRNA, generating an active TF (GmbZIP68S) lacking the transmembrane segment. GmbZIP68S (AtbZIP60S) moves to the nucleus to induce the expression of molecular chaperones, ERAD components,

GmbZIP68 (AtbZIP60), GmNAC021 (AtNAC062) and GmNAC103 (AtNAC089). Evidence indicates that GmbZIP37/38 and GmbZIP68S may act in concert as heterodimers to activate ER stress-responsive genes. GmbZIP68S (AtbZIP60S) also induces expression of the TF GmNAC020 (AtNAC103) to further amplify the ER stress response (19). As a plasma membrane component of the ER stress response (17), membrane-tethered GmNAC021 (AtNAC62) also undergoes regulated intramembrane proteolysis (RIP) for release into the nucleus as a positive regulator of ER stress-responsive genes. If the UPR is not capable of restoring ER homeostasis under prolonged and severe stress, then PCD responses are activated for the regulated disposal of abnormal cells. ER stress-induced proteolysis of ER membrane-tethered GmNAC103 (AtNAC089) exemplifies an ER stress-induced plant-specific PCD response (20). RIP-mediated translocation of GmNAC103 to the nucleus allows for the induction of PCD-associated gene expression, promoting DNA fragmentation and an increase in caspase-3/7-like activity. A distinct ER stress-induced PCD response in soybean integrates an osmotic stress signal into a full response (22). The combination of ER stress and osmotic stress fully induces the expression of the TF GmERD15 (22) to activate the expression of the membrane-associated protein DCD/NRP-B (23). Induction of DCD/NRP-B activates a signaling cascade that culminates with the induction of the GmNAC081 and GmNAC030 TFs (24), which form heterodimers to fully transactivate the vacuolar processing enzyme (VPE) promoter (25). VPE exhibits caspase-1-like activity and induces plant-specific PCD, mediated by collapse of the vacuole (26).

TABLES

Table 1 - Copy numbers of UPR genes

Arabidopsis designation	Gene copy number in Arabidopsis	Gene copy number in soybean
Transducers/sensors		
bZIP17	1	-
bZIP18	1	-
bZIP17/28	-	2
IRE1A	1	3
IRE1B	1	1
Immediate downstream components		
bZIP60	1	1
NAC103	1	4
S1P	1	2
S2P	1	1
SAR1	2	10
Sec12	3	1
Downstream components		
1) Molecular chaperones/foldases		
BiP	3	4
CRT	3	6
CNX	4	11
PDI	13	22
PPI	1	2
Erdj3	3	8
GRP54	1	2
2) Folding of glycoproteins		
OST	2	4
Glc-I	1	3
Glc-II	3	5
UGGT	1	3
3) ERAD		
MNS3	1	2
MNS4	1	1
MNS5	1	2
EBS6/OS9	1	2
EBS5/HRD3A	1	2
HRD3B	1	Not found
HRD1A-HRD1B	2	2
Der	3	4
PUXUBX2	15	17
UBC32	1	4
UBC33/34	2	2
CDC48	3	10
UFD1	4	6
NPL4	2	2
UDF2	1	2
PNG1	1	2
RAD23	4	7
DSK2	2	4
ER stress/plasma membrane cooperative response		
NAC62	1	3

Table 2 – Copy numbers of genes involved in ER stress-induced PCD

1) DCD/NRP-mediated cell death response		
Soybean designation	Gene copy number in soybean	Gene copy number in Arabidopsis
GmERD15	2	No description
DCD/NRP-A and B	4 (2 NRP-A and 2 NRP-B)	1
DCD/NRP-C	2	1
NAC81	2	1
NAC30	7	1
VPE	4	2 (alpha-VPE and gamma-VPE)
2) NAC89-mediated cell death signaling		
Arabidopsis designation	Gene copy number in Arabidopsis	Gene copy number in soybean
NAC89	1	2
3) AGB1-mediated cell death signaling		
Arabidopsis designation	Gene copy number in Arabidopsis	Gene copy number in soybean
AGB1	1	4

ADDITIONAL FILES

Addirional file 1. A list of known/predicted orthologous genes involved in the *Glycine max* and *Arabidopsis thaliana* UPR pathways.

Transducers/sensors											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
bZIP 17/28	Glyma.03G123200.1.p	bZIP transcription factor (bZIP38)	virNOG01396	AT2G40950.1 AT3G10800.1 AT3G56660.1	bZIP17 bZIP28 bZIP49	AT2G40950.1	8e-171	48,03	60,66	96	19, 26, 20
bZIP 17/28	Glyma.19G126800.1.p	bZIP transcription factor (bZIP37)	virNOG01396	AT2G40950.1 AT3G10800.1 AT3G56660.1	bZIP17 bZIP28 bZIP49	AT2G40950.1	2e-167	47,68	61,32	94	
IRE (Template: AT2G17520.1)	Glyma.01G157800.1.p	IRE1P-Related	virNOG09069	AT2G17520.1 AT5G24360.2	ATIRE1-A IRE1B	AT2G17520.1	0	68,86	82,27	70	12, 14
IRE	Glyma.09G197000.1.p	IRE1P-Related	virNOG09069	AT2G17520.1 AT5G24360.2	ATIRE1-A IRE1B	AT2G17520.1	0	66,59	81,94	49	
IRE	Glyma.11G087200.1.p	IRE1P-Related	virNOG09069	AT2G17520.1 AT5G24360.2	ATIRE1-A IRE1B	AT2G17520.1	0	65,73	80,17	74	
IRE	Glyma.16G111800.1.p	IRE1P-Related	virNOG09069	AT2G17520.1 AT5G24360.2	ATIRE1-A IRE1B	AT5G24360.2	0	43,14	60,67	94	
Immediate dowstream components											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
bZIP 60 (Template: AT1G42990.1)	Glyma.02G161100.1.p	GmbZIP68 (Nomenclature Liao et al. 2008)	euNOG19243	AT1G42990.1	ATBZIP60	AT1G42990.1	3,00E-020	46,62	57,43	52	50, 14
NAC103 (Template: AT5G64060)	Glyma.04G213300.1.p	GmNAC020 (Nomenclature Le et. al, 2011)	virNOG18312	AT5G09330.1 AT5G64060.1	ANAC082 ANAC103	AT5G64060.1	3,00E-087	52,32	70,04	49	
NAC103	Glyma.05G191300.1.p	GmNAC028 (Nomenclature Le et. al, 2011)	virNOG18312	AT5G09330.1 AT5G64060.1	ANAC082 ANAC103	AT5G09330.1	4,00E-067	40,97	56,25	76	31
NAC103	Glyma.08G156500.1.p	GmNAC058 (Nomenclature Le et. al, 2011)	virNOG18312	AT5G09330.1 AT5G64060.1	ANAC082 ANAC103	AT5G09330.1	2,00E-083	44,31	58,6	96	
NAC103	Glyma.06G152900.1.p	GmNAC037 (Nomenclature Le et. al, 2011)	virNOG18312	AT5G09330.1 AT5G64060.1	ANAC082 ANAC103	AT5G09330.1	3,00E-090	52,57	69,96	50	
S1P (Template: AT5G19660.1)	Glyma.07G085500.1.p	Site-1 protease.	virNOG03101	AT5G19660.1	ATS1P	AT5G19660.1	0	75,08	84,58	96	45
S1P	Glyma.09G191400.1.p	Site-1 protease.	virNOG03101	AT5G19660.1	ATS1P	AT5G19660.1	0	73,03	82,56	99	45

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
SAR1 (Template: AT1G56330.1 and AT4G02080)	Glyma.03G231800.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	1e-139	93,26	99,48	99	46
SAR1	Glyma.07G241800.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	2e-140	93,26	98,96	99	46
SAR1	Glyma.07G241900.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	1e-139	92,23	99,48	99	46
SAR1	Glyma.09G038200.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	1e-141	93,78	99,48	99	46
SAR1	Glyma.10G147800.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	2e-138	91,71	98,45	99	46
SAR1	Glyma.15G143300.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	2e-141	93,78	99,48	99	46
SAR1	Glyma.17G032000.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	1e-138	91,71	99,48	99	46
SAR1	Glyma.17G032100.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	5e-140	92,75	98,96	99	46
SAR1	Glyma.19G228800.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	7e-140	93,78	98,96	99	46
SAR1	Glyma.20G239600.1.p	Vesicle coat complex COPII, GTPase subunit SAR1	EggNOG-KOG0077	AT1G56330.1 AT4G02080	SAR1-B SAR1	AT4G02080.1	6e-138	91,71	97,93	99	46
Sec12 (Template: AT2G01470.1)	Glyma.07G218700.1.p	Prolactin regulatory element-binding protein/Protein transport protein SEC12p	virNOG01500	AT2G01470.1 AT5G50550.1 AT5G50650.1	SEC12P-LIKE 2	AT5G50650.1	1e-112	48,04	67,62	95	
Dowstream components											
1) Molecular chaperones/foldases											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
BiP (Template: AT5G28540.1 AtBiP1)	Glyma.05G219600.1.p	Molecular chaperones GRP78/BiP/KAR2, HSP70 superfamily	virNOG09258	AT5G28540.1 AT5G42020.1	BiP1 BiP2	AT5G42020.1	0	90,88	95,37	95	

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
BiP	Glyma.05G219400.1.p	Molecular chaperones GRP78/BiP/KAR2, HSP70 superfamily - soyBiP D e soyBiP B	virNOG09258	AT5G28540.1 AT5G42020.1	BiP1 BiP2	AT5G42020.1	0	91,78	96,11	100	59
BiP	Glyma.08G025700.1.p	Molecular chaperones GRP78/BiP/KAR2, HSP70 superfamily - soyBiP C	virNOG09258	AT5G28540.1 AT5G42020.1	BiP1 BiP2	AT5G42020.1	0	91,78	95,81	100	59, 61
BiP	Glyma.08G025900.1.p	Molecular chaperones GRP78/BiP/KAR2, HSP70 superfamily - soyBiP A	virNOG09258	AT5G28540.1 AT5G42020.1	BiP1 BiP2	AT5G42020.1	0	90,45	95,52	94	59
BiP	not found		virNOG29237	AT1G09080.1	BIP3	BIP3					
CRT (Template: AT1G08450.1)	Glyma.09G248600.1.p	Calreticulin	virNOG02900	AT1G08450.1	AtCRT3	AT1G08450.1	0	72,97	83,25	90	32, 67
CRT	Glyma.11G126100.1.p	Calreticulin	virNOG02900	AT1G08450.1	AtCRT3	AT1G08450.1	0	79,05	89,05	99	
CRT	Glyma.12G050700.1.p	Calreticulin	virNOG02900	AT1G08450.1	AtCRT3	AT1G08450.1	0	80,63	91,04	98	
CRT	Glyma.18G244100.1.p	Calreticulin	virNOG02900	AT1G08450.1	AtCRT3	AT1G08450.1	0	68,09	78,01	78	
CRT (Template: AT1G09210.1 and AT1G56340.1)	Glyma.10G147600.1.p	Calreticulin	virNOG10578	AT1G09210.1 AT1G56340.1	AtCRT1b AtCRT1a	AT1G09210.1	0	86,1	92,51	86	
CRT	Glyma.20G098400.1.p	Calreticulin	virNOG10578	AT1G09210.1 AT1G56340.1	AtCRT1b AtCRT1a	AT1G09210.1	0	85,51	92,9	78	
CNX (Template: AT5G55130.1)	Glyma.02G293600.1.p	Molybdenum Cofactor Synthesis Protein	virNOG06123	AT5G55130.1	CNX5	AT5G55130.1	0	68,91	82,61	99	32
CNX	Glyma.14G019100.1.p	Molybdenum Cofactor Synthesis Protein	virNOG06123	AT5G55130.1	CNX5	AT5G55130.1	0	68,04	83,26	99	
CNX	Glyma.14G020700.1.p	Molybdenum Cofactor Synthesis Protein	virNOG06123	AT5G55130.1	CNX5	AT5G55130.1	0	72,57	85,14	99	
CNX (Template: AT1G01290.1)	Glyma.20G102700.1.p	Molybdopterin Cofactor Synthesis Protein A	virNOG06264	AT1G01290.1	CNX3	AT1G01290.1	5E-096	57,53	68,73	97	
CNX (Template: AT2G31955.1)	Glyma.08G157800.1.p	Molybdopterin Cofactor Synthesis Protein A	virNOG09352	AT2G31955.1	CNX2	AT2G31955.1	0	72,82	85,29	100	

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
CNX	Glyma.15G268400.1.p	Molybdopterin Cofactor Synthesis Protein A	virNOG09352	AT2G31955.1	CNX2	AT2G31955.1	0	74,31	86,53	100	
CNX	Glyma.06G321300.1.p	Molybdopterin converting factor, small subunit	virNOG23307	AT4G10100.1	CNX7	AT4G10100.1	8E-042	61,17	77,67	100	
CNX (Template: AT5G61790.1)	Glyma.04G202900.1.p	Calnexin	virNOG13792	AT5G61790.1	ATCNX1	AT5G61790.1	0	71,05	81,23	98	
CNX	Glyma.05G199200.1.p	Calnexin	virNOG13792	AT5G61790.1	ATCNX1	AT5G61790.1	0	73,87	85,53	98	
CNX	Glyma.06G162600.1.p	Calnexin	virNOG13792	AT5G61790.1	ATCNX1	AT5G61790.1	0	77,96	88,37	89	
CNX	Glyma.08G006500.1.p	Calnexin	virNOG13792	AT5G61790.1	ATCNX1	AT5G61790.1	0	74,12	84,92	99	
PDI (Template: AT2G47470.1)	Glyma.19G229000.1.p	GmPDI-S3 (Nomenclature Selles et al. 2011)	virNOG09353	AT2G47470.1	PDI 2.1 (Nomenclature Houston et al. 2005)	AT2G47470.1	3E-178	70,91	82,73	86	68-70
PDI	Glyma.03G232000.1.p	GmPDI-S4 (Nomenclature Selles et al. 2011)	virNOG09353	AT2G47470.1	PDI 2.1 (Nomenclature Houston et al. 2005)	AT2G47470.1	0	75,45	86,36	91	68-70
PDI	Glyma.02G014000.1.p	GmPDI-S1 (Nomenclature Selles et al. 2011)	virNOG09353	AT2G47470.1	PDI 2.1 (Nomenclature Houston et al. 2005)	AT2G47470.1	0	77,68	88,1	91	68-70
PDI	Glyma.10G014700.1.p	GmPDI-S2 (Nomenclature Selles et al. 2011)	virNOG09353	AT2G47470.1	PDI 2.1 (Nomenclature Houston et al. 2005)	AT2G47470.1	0	78,92	88,86	91	68-70
PDI (Template: AT3G20560.1)	Glyma.12G213100.1.p	GmPDI-C2 (Nomenclature Selles et al. 2011)	virNOG04036	AT3G20560.1 AT4G27080.2	PDI 5.3 PDI 5.4 (Nomenclature Houston et al. 2005)	AT3G20560.1	0	72,73	86,16	100	68-70
PDI	Glyma.13G288600.1.p	GmPDI-C1 (Nomenclature Selles et al. 2011)	virNOG04036	AT3G20560.1 AT4G27080.2	PDI 5.3 PDI 5.4 (Nomenclature Houston et al. 2005)	AT3G20560.1	0	71,69	85,95	100	68-70

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
PDI	Glyma.06G270200.1.p	GmPDI-C4 (Nomenclature Selles et al. 2011)	virNOG04036	AT3G20560.1 AT4G27080.2	PDI 5.3 PDI 5.4 (Nomenclature Houston et al. 2005)	AT4G27080.2	0	74,57	86,11	97	68-70
PDI (Template: AT1G04980.1)	Glyma.02G266900.1.p	GmPDI-M2 (Nomenclature Selles et al. 2011)	virNOG01952	AT1G04980.1 AT2G32920.1	PDI 2.2 PDI 2.3 (Nomenclature Houston et al. 2005)	AT1G04980.1	0	73,56	85,11	100	68-70
PDI	Glyma.14G050600.1.p	GmPDI-M1 (Nomenclature Selles et al. 2011)	virNOG01952	AT1G04980.1 AT2G32920.1	PDI 2.2 PDI 2.3 (Nomenclature Houston et al. 2005)	AT1G04980.1	0	72,67	84,44	100	68-70
PDI (Template: AT1G77510.1)	Glyma.13G077300.1.p	GmPDI-L1c (Nomenclature Selles et al. 2011)	virNOG06651	AT1G21750.1 AT1G77510.1	PDI 1.1 PDI 1.2 (Nomenclature Houston et al. 2005)	AT1G21750.1	0	67,71	83,23	95	68-70
PDI	Glyma.14G152000.1.p	Protein disulfide isomerase	virNOG06651	AT1G21750.1 AT1G77510.1	PDI 1.1 PDI 1.2 (Nomenclature Houston et al. 2005)	AT1G21750.1	0	69,47	83,58	95	68-70
PDI	Glyma.04G247900.1.p	GmPDI-L1a (Nomenclature Selles et al. 2011)	virNOG06651	AT1G21750.1 AT1G77510.1	PDI 1.1 PDI 1.2 (Nomenclature Houston et al. 2005)	AT1G77510.1	0	70,04	82,85	92	68-70
PDI	Glyma.06G114800.1.p	GmPDI-L1b (Nomenclature Selles et al. 2011)	virNOG06651	AT1G21750.1 AT1G77510.1	PDI 1.1 PDI 1.2 Nomenclature Houston et al. 2005)	AT1G77510.1	0	69,57	81,99	94	68-70

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
PDI (Template: AT1G35620.1)	Glyma.10G217600.1.p	GmPDI-B (Nomenclature Selles et al. 2011)	virNOG01884	AT1G35620.1	PDI 5.2 (Nomenclature Houston et al . 2005)	AT1G35620.1	0	56,53	76,13	99	68-70
PDI (Template: AT3G54960.1)	Glyma.12G172800.1.p	GmPDI-L2d (Nomenclature Selles et al. 2011)	virNOG06454	AT3G54960.1 AT5G60640.1	PDI 1.3 PDI 1.4 (Nomenclature Houston et al. 2005)	AT5G60640.1	0	70,34	87,37	82	68-70
PDI	Glyma.13G326200.1.p	GmPDI-L2c (Nomenclature Selles et al. 2011)	virNOG06454	AT3G54960.1 AT5G60640.1	PDI 1.3 PDI 1.4 (Nomenclature Houston et al. 2005)	AT5G60640.1	0	70,02	85,69	90	68-70
PDI (Template: AT5G60640.1)	Glyma.11G146800.1.p	GmPDI-L2a (Nomenclature Selles et al. 2011)	eggNOG-KOG0190	AT5G60640.1	PDI 1.4 (Nomenclature Houston et al. 2005)	AT5G60640.1	0,00E+000	70,21	83,95	86	68-70
PDI	Glyma.12G067700.1.p	GmPDI-L2b (Nomenclature Selles et al. 2011)	eggNOG-KOG0190	AT5G60640.1	PDI 1.4 (Nomenclature Houston et al. 2005)	AT5G60640.1	0,00E+000	69,75	83,43	87	68-70
PDI (Template: AT1G52260.1)	Glyma.13G357700.1.p	GmPDI-L3a (Nomenclature Selles et al. 2011)	virNOG00048	AT1G52260.1 AT3G16110.1	PDI 1.5 PDI 1.6 (Nomenclature Houston et al. 2005)	AT3G16110.1	0	58,79	76,97	95	68-70
PDI	Glyma.15G015700.1.p	GmPDI-L3b (Nomenclature Selles et al. 2011)	virNOG00048	AT1G52260.1 AT3G16110.1	PDI 1.5 PDI 1.6 (Nomenclature Houston et al. 2005)	AT3G16110.1	0	62,42	82,63	94	68-70
PDI (Template: AT1G07960.1)	Glyma.13G328300.1.p	GmPDI-A2 (Nomenclature Selles et al. 2011)	virNOG20797	AT1G07960.1	PDI 5.1 (Nomenclature Houston et al. 2005)	AT1G07960.1	2,00E-073	69,78	84,89	92	68-70
PDI	Glyma.15G045500.1.p	GmPDI-A1 (Nomenclature Selles et al. 2011)	virNOG20797	AT1G07960.1	PDI 5.1 (Nomenclature Houston et al. 2005)	AT1G07960.1	7,00E-074	67,36	83,33	99	68-70

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
PIN1AT/PPIase (Template AT1G18040)	Glyma.02G049500.1.p	peptidylprolyl cis/trans isomerase, NIMA-interacting 1	virNOG21280	AT2G18040	PIN1AT	AT2G18040	3e-71	82,3	94,69	89	71, 72
PIN1AT/PPIase	Glyma.16G129900.1.p	peptidylprolyl cis/trans isomerase, NIMA-interacting 1	virNOG21280	AT2G18040	PIN1AT	AT2G18040	1e-71	83,19	95,58	89	
ERdj1(Template AT1G28210)	Glyma.02G211200.1.p	DNAJ heat shock family protein	virNOG02787	AT1G28210.2	ATJ1	AT1G28210.2	5,00E-179	61,64	75,45	83	73
ERdj1	Glyma.14G178800.1.p	DNAJ heat shock family protein	virNOG02787	AT1G28210.2	ATJ1	AT1G28210.2	0,00E+000	62,15	76,21	83	
ERdj2 (Template AT5G22060)	Glyma.03G116600.1.p	J2 DNAJ homologue 2	virNOG00916	AT5G22060.1	ATJ2	AT5G22060.1	0,00E+000	70,21	83,22	100	73
ERdj2	Glyma.07G110200.1.p	J2 DNAJ homologue 2	virNOG00916	AT5G22060.1	ATJ2	AT5G22060.1	0,00E+000	69,74	82,98	100	
ERdj2	Glyma.U012100.1.p	J2 DNAJ homologue 2	virNOG00916	AT5G22060.1	ATJ2	AT5G22060.1	0,00E+000	82,7	90,76	100	
ERdj2	Glyma.12G095700.1.p	J2 DNAJ homologue 2	virNOG00916	AT5G22060.1	ATJ2	AT5G22060.1	0,00E+000	83,18	91,47	100	
ERdj3 (Template AT3G44110)	Glyma.13G311600.1.p	J2 DNAJ homologue 2	EggNOG-KOG0712	AT3G44110	ATJ3	AT3G44110	0,00E+000	83,65	90,52	100	73
ERdj3	Glyma.12G190100.1.p	J2 DNAJ homologue 2	EggNOG-KOG0712	AT3G44110	ATJ3	AT3G44110	0,00E+000	83,18	90,28	100	
GRP94 (Template AT4G24190)	Glyma.14G219700.1.p	SHD Chaperone protein htpG family protein	virNOG04777	AT4G24190	ATHSP90.7	AT4G24190	0,00E+000	79,63	87,78	95	74
GRP94	Glyma.17G258700.1.p	SHD Chaperone protein htpG family protein	virNOG04777	AT4G24190	ATHSP90.7	AT4G24190	0,00E+000	82,67	91,27	95	
2) Folding of glycoproteins											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
OST (Template: AT4G29870.1)	Glyma.09G061600.1.p	OST3 / OST6 family	virNOG17351	AT4G29870.1		AT4G29870.1	8,00E-099	85,09	94,41	91	7, 75
OST	Glyma.15G167900.1.p	OST3 / OST6 family	virNOG17351	AT4G29870.1		AT4G29870.1	1,00E-096	83,44	94,27	89	
OST (Template: AT1G61790.1)	Glyma.20G141900.1.p	OST3 / OST6 family	virNOG03318	AT1G61790.1		AT1G61790.1	2,00E-161	67,2	83,6	92	
OST	Glyma.10G251600.1.p	OST3 / OST6 family	virNOG03318	AT1G61790.1		AT1G61790.1	4,00E-161	68,17	83,28	92	
Glc-I (Template: AT1G67490.1)	Glyma.05G146500.1.p	Mannosyl-Oligosaccharide Glucosidase	virNOG09226	AT1G67490.1	GCS1	AT1G67490.1	0	61,55	76,44	99	7, 76

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
Glc-I	Glyma.08G103200.1.p	Mannosyl-Oligosaccharide Glucosidase	virNOG09226	AT1G67490.1	GCS1	AT1G67490.1	0	57,24	70,68	99	
Glc-I	Glyma.05G146400.1.p	Mannosyl-Oligosaccharide Glucosidase	virNOG09226	AT1G67490.1	GCS1	AT1G67490.1	0	64,18	77,96	93	
Glc-II (Template: AT2G42390.1)	Glyma.12G120800.1.p	Glucosidase II BETA Subunit	virNOG20019	AT2G42390.1		AT2G42390.1	1E-068	55,68	66,49	74	77
Glc-II (Template: AT5G56360.1)	Glyma.07G224600.1.p	Glucosidase II BETA Subunit	virNOG07907	AT5G56360.1	PSL4	AT5G56360.1	0	59,62	72,81	98	
Glc-II	Glyma.20G024500.1.p	Glucosidase II BETA Subunit	virNOG07907	AT5G56360.1	PSL4	AT5G56360.1	0	58,2	72,4	97	
Glc-II (Template: AT5G63840.1)	Glyma.04G209000.1.p	Glucosidase II catalytic (alpha) subunit and related enzymes	virNOG06643	AT5G63840.1	PSL5	AT5G63840.1	0	73,64	84,16	99	
Glc-II	Glyma.06G156900.1.p	Glucosidase II catalytic (alpha) subunit and related enzymes	virNOG06643	AT5G63840.1	PSL5	AT5G63840.1	0	72,08	82,55	98	
UGGT/EBS1 (Template: AT1G71220.2)	Glyma.05G168600.1.p	UDP-glucose:Glycoprotein Glucosyltransferase	virNOG03272	AT1G71220.2	EBS1	AT1G71220.2	0	65,24	79,12	100	7, 66, 78
UGGT/EBS1	Glyma.08G127200.1.p	UDP-glucose:Glycoprotein Glucosyltransferase	virNOG03272	AT1G71220.2	EBS1	AT1G71220.2	0	65,66	78,98	100	7, 66, 78
UGGT/EBS1	Glyma.08G187500.1.p	UDP-glucose:Glycoprotein Glucosyltransferase	virNOG03272	AT1G71220.2	EBS1	AT1G71220.2	0	68,67	81,48	99	7, 66, 78
3) ERAD											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
MNS3 (Template: AT1G30000.1)	Glyma.13G217900.1.p	Mannosyl-Oligosaccharide ALPHA-1,2-Mannosidase-Related	virNOG05013	AT1G30000.1	MNS3	AT1G30000.1	0	72,59	81,93	99	80
MNS3	Glyma.15G094500.1.p	Mannosyl-Oligosaccharide ALPHA-1,2-Mannosidase-Related	virNOG05013	AT1G30000.1	MNS3	AT1G30000.1	0	73,52	82,24	99	80

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
MNS4 (Template: AT5G43710.1)	Glyma.10G179900.1.p	Glycosyl hydrolase family 47 protein	virNOG04591	AT5G43710.1		AT5G43710.1	0	73,82	84,12	95	80, 84
MNS5 (Template: AT1G27520.1)	Glyma.05G034300.1.p	Glycosyl hydrolase family 47 protein	virNOG16095	AT1G27520.1		AT1G27520.1	2E-100	72,31	82,56	96	80, 84
MNS5	Glyma.17G092900.2.p	Glycosyl hydrolase family 47 protein	virNOG16095	AT1G27520.1		AT1G27520.1	0	77,39	89,28	98	80, 84
EBS6/OS9 (Template: AT5G35080.1)	Glyma.09G091600.1.p	OS-9-Related	virNOG03239	AT5G35080.1	AtOS9	AT5G35080.1	5E-118	58,55	73,09	91	80, 86
EBS6/OS9 (Template: AT5G35080.1)	Glyma.15G198100.1.p	OS-9-Related	virNOG03239	AT5G35080.1	AtOS9	AT5G35080.1	5E-119	61,36	74,62	88	80, 86
EBS5 (Template: AT1G18260.1)	Glyma.05G246700.2.p	HCP-like superfamily protein	virNOG05616	AT1G18260.1	EBS5	AT1G18260.1	0	72,76	85,14	94	85, 86
EBS5/HRD3A	Glyma.08G054900.2.p	HCP-like superfamily protein	virNOG05616	AT1G18260.1	EBS5	AT1G18260.1	0	72,64	85,01	84	85, 86
HRD3B			KOG1550	AT1G73570.1							85, 86
HRD1A-HRD1B (Template: AT3G16090.1)	Glyma.01G228500.2.p	RING/U-box superfamily protein	virNOG07581	AT1G65040.2 AT3G16090.1	AtHrd1B AtHrd1A	AT3G16090.1	0	65,1	76,17	97	85, 86
HRD1A-HRD1B	Glyma.11G011300.1.p	RING/U-box superfamily protein	virNOG07581	AT1G65040.2 AT3G16090.1	AtHrd1B AtHrd1A	AT3G16090.1	0	64,76	75,28	97	85, 86
DER1 (Template: AT4G29330.1)	Glyma.05G225500.1.p	Derlin-1	virNOG01130	AT4G29330.1	DER1	AT4G29330.1	6E-125	68,05	74,81	99	33
DER1	Glyma.08G032300.1.p	Derlin-1	virNOG01130	AT4G29330.1	DER1	AT4G29330.1	1E-124	67,67	74,44	99	33
DER2.1-DER2.2 (Template: AT4G04860.1)	Glyma.13G205900.1.p	Derlin-2.2	virNOG02714	AT4G04860.1 AT4G21810.1	DER2.2 DER2.1	AT4G04860.1	1E-170	89,39	96,33	100	32, 91
DER2.1-DER2.2	Glyma.15G106600.1.p	Derlin-2.2	virNOG02714	AT4G04860.1 AT4G21810.1	DER2.2 DER2.1	AT4G04860.1	1E-170	89,39	96,33	100	32, 91
Usa1*				Not found							34
PUX1 (Template: AT3G27310.1)	Glyma.16G063100.1.p	Plant UBX domain-containing protein 1	virNOG01649	AT3G27310.1	PUX1	AT3G27310.1	8E-100	59,39	72,03	100	95, 96
PUX1	Glyma.19G082400.1.p	Plant UBX domain-containing protein 1	virNOG01649	AT3G27310.1	PUX1	AT3G27310.1	5E-100	62,3	74,6	97	95, 96
PUX2	Glyma.05G235400.1.p	Plant UBX domain-containing protein 2	virNOG06034	AT2G01650.1	PUX2	AT2G01650.1	0	59,66	73,11	99	108
PUX2	Glyma.08G042900.1.p	Plant UBX domain-containing protein 2	virNOG06034	AT2G01650.1	PUX2	AT2G01650.1	0	59,47	72,84	99	108

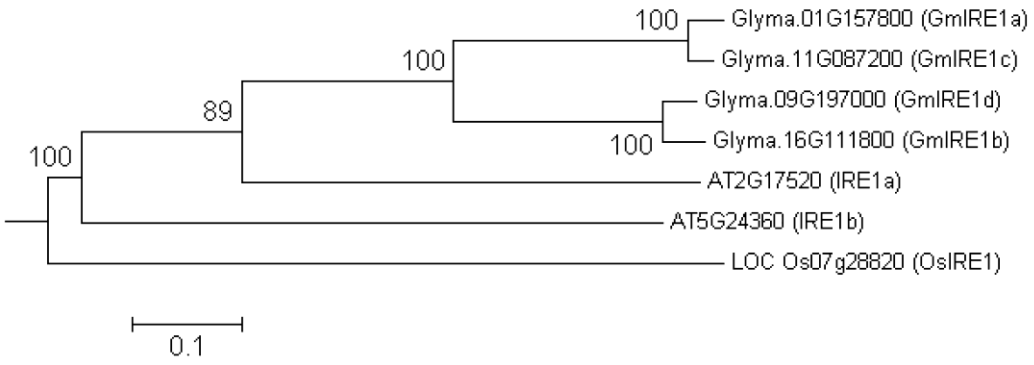
Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
PUX3-PUX4-PUX5 (Template: AT4G04210.1)	Glyma.02G270000.1.p	Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B prime gamma	virNOG03930	AT4G04210.1 AT4G15410.1 AT4G22150.1	PUX4 PUX5 PUX3	AT4G15410.1	1E-138	54,69	66,43	98	108
PUX3-PUX4-PUX5	Glyma.09G007600.1.p	Plant UBX domain containing protein 4	virNOG03930	AT4G04210.1 AT4G15410.1 AT4G22150.1	PUX4 PUX5 PUX3	AT4G04210.1	8E-141	70,1	79,74	100	108
PUX3-PUX4-PUX5	Glyma.14G046900.1.p	Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B prime gamma	virNOG03930	AT4G04210.1 AT4G15410.1 AT4G22150.1	PUX4 PUX5 PUX3	AT4G15410.1	1E-137	53,36	64,97	98	108
PUX3-PUX4-PUX5	Glyma.15G112000.1.p	Plant UBX domain containing protein 4	virNOG03930	AT4G04210.1 AT4G15410.1 AT4G22150.1	PUX4 PUX5 PUX3	AT4G04210.1	1E-141	70,74	80,39	100	108
PUX6			KOG2086	AT3G21660							108
PUX7 (Template: AT1G14570.1)	Glyma.08G267900.1.p	UBX domain-containing protein	virNOG11065	AT1G14570.1	AtPUX7	AT1G14570.1	0	58,81	71,72	100	108
PUX7	Glyma.16G100500.1.p	UBX domain-containing protein	virNOG11065	AT1G14570.1	AtPUX7	AT1G14570.1	0	58,13	71,75	100	108
PUX8/SAY1 (Template: AT4G11740.1)	Glyma.05G049200.1.p	Ubiquitin-like superfamily protein	virNOG03808	AT4G11740.1 AT4G23040.1	PUX8 / SAY1 PUX13	AT4G11740.1	5E-151	49,35	62,62	100	108
PUX8/SAY1	Glyma.17G131100.1.p	Ubiquitin-like superfamily protein	virNOG03808	AT4G11740.1 AT4G23040.1	PUX8 / SAY1 PUX13	AT4G11740.1	6E-155	49,51	64,12	100	108
PUX9 (Template: AT4G00752.1)	Glyma.07G077200.1.p	Ubiquitin-like superfamily protein	virNOG16991	AT4G00752.1	PUX9	AT4G00752.1	1e-67	35,07	51,06	99	108
PUX10 (Template: AT4G10790.1)	Glyma.10G173300.1.p	UBX domain-containing protein	virNOG05342	AT4G10790.1	PUX10	AT4G10790.1	0	69,71	80,08	99	108
PUX10	Glyma.20G216900.1.p	UBX domain-containing protein	virNOG05342	AT4G10790.1	PUX10	AT4G10790.1	0	65,62	76,46	99	108
PUX11			virNOG11675	AT2G43210.1	PUX11						108
PUX12			virNOG35532	AT3G23605.1	PUX12						108
PUX14			virNOG29703	AT4G14250.1	PUX14						108
PUX15			KOG1364	AT1G59550							108
D0A10 (Template: AT4G34100.1)	Glyma.02G103800.1.p	RING/U-box superfamily protein	virNOG06272	AT4G34100.1	CER9	AT4G34100.1	0	75,09	85,2	99	18, 87

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
D0A10	Glyma.07G215200.1.p	RING/U-box superfamily protein	virNOG06272	AT4G34100.1	CER9	AT4G34100.1	0	77,22	87,13	93	18, 87
Cue1*				Not found							34
UBC32 (Template: AT3G17000.1)	Glyma.07G240600.1.p	Ubiquitin-conjugating enzyme 32	virNOG06265	AT3G17000.1	UBC32	AT3G17000.1	3E-134	61,83	74,45	99	88, 90
UBC32	Glyma.09G036300.1.p	Ubiquitin-conjugating enzyme 32	virNOG06265	AT3G17000.1	UBC32	AT3G17000.1	3E-137	64,33	74,84	99	88, 90
UBC32	Glyma.15G141200.1.p	Ubiquitin-conjugating enzyme 32	virNOG06265	AT3G17000.1	UBC32	AT3G17000.1	3E-138	66,02	75,4	99	88, 90
UBC32	Glyma.17G032800.1.p	Ubiquitin-conjugating enzyme 32	virNOG06265	AT3G17000.1	UBC32	AT3G17000.1	5E-137	61,8	74,22	99	88, 90
UBC33-UBC34 (Template: AT1G17280)	Glyma.07G046200.1.p	Ubiquitin-conjugating enzyme 34	virNOG00673	AT1G17280.1 AT5G50430.1	UBC34 UBC33	AT1G17280.1	2E-139	78,84	85,89	100	88, 90
UBC33-UBC34	Glyma.16G014400.1.p	Ubiquitin-conjugating enzyme 34	virNOG00673	AT1G17280.1 AT5G50430.1	UBC34 UBC33	AT1G17280.1	9e-138	78,84	85,89	100	88, 90
CDC48A (Template: AT3G09840.1)	Glyma.03G182800.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	92,62	97,12	99	94
CDC48A	Glyma.04G186000.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	87,95	94,86	97	94
CDC48A	Glyma.06G180000.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	88,52	94,96	97	94
CDC48A	Glyma.10G057100.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	92,15	96,64	99	94
CDC48A	Glyma.12G177100.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	91,75	96,31	100	94
CDC48A	Glyma.13G143600.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	92,27	96,63	99	94
CDC48A	Glyma.13G323600.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	91,63	96,31	100	94

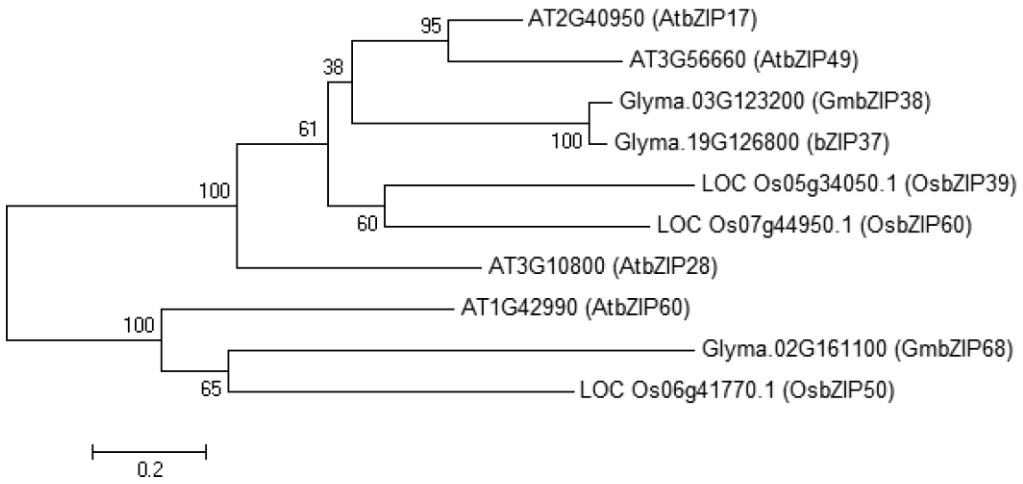
Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
CDC48A	Glyma.19G183400.1.p	ATPase, AAA-type, CDC48 protein	virNOG05493	AT3G09840.1 AT3G53230.1 AT5G03340.1	CDC48A CDC48 CDC48	AT5G03340.1	0	93,13	97,13	99	94
CDC48B	Glyma.02G118800.1.p	Cell division cycle 48B	virNOG11486	AT2G03670.1	CDC48B	AT2G03670.1	0	77,92	88,32	90	94
CDC48C	Glyma.08G186800.1.p	Cell division cycle 48C	virNOG08944	AT3G01610.1	CDC48C	AT3G01610.1	0	57,76	71,5	97	94
UFD1 (Template: AT2G21270.3)	Glyma.05G238500.1.p	Ubiquitin fusion degradation 1	virNOG00505	AT2G21270.3 AT4G38930.2	UFD1 UFD1	AT2G21270.3	1e-171	75,93	84,88	100	109
UFD1	Glyma.08G045500.1.p	Ubiquitin fusion degradation 1	virNOG00505	AT2G21270.3 AT4G38930.2	UFD1 UFD1	AT2G21270.3	3E-173	75,78	83,85	100	109
UFD1 (Template:AT2G29070.2)	Glyma.13G337800.1.p	Ubiquitin Fusion Degradaton Protein 1	virNOG07963	AT2G29070.2	UFD1	AT2G29070.2	1E-115	56,37	68,47	99	109
UFD1	Glyma.15G036700.1.p	Ubiquitin Fusion Degradaton Protein 1	virNOG07963	AT2G29070.2	UFD1	AT2G29070.2	1E-149	65,3	77,29	98	109
UFD1	Glyma.02G271100.1.p	Ubiquitin fusion degradation UFD1 family protein	virNOG01917	AT4G15420.1	UFD1	AT4G15420.1	0	65,68	82,75	99	109
UFD1	Glyma.14G046100.1.p	Ubiquitin fusion degradation UFD1 family protein	virNOG01917	AT4G15420.1	UFD1	AT4G15420.1	0	65,21	81,29	99	109
NPL4 (Template: AT2G47970)	Glyma.03G207900.1.p	NPL4-like protein 1	virNOG02773	AT2G47970.1 AT3G63000.1	NPL4 NPL41	AT3G63000.1	0	73,91	87,2	99	
NPL4	Glyma.19G205200.1.p	NPL4-like protein 1	virNOG02773	AT2G47970.1 AT3G63000.1	NPL4 NPL41	AT3G63000.1	0	74,15	87,2	99	
UFD3*				Not found							34
UFD2 (Template: AT5G15400.1)	Glyma.13G059200.2.p	U-box domain-containing protein	virNOG08159	AT5G15400.1	U-box domain-containing protein	AT5G15400.1	0	75,98	86,18	99	
UFD2	Glyma.19G027200.2.p	U-box domain-containing protein	virNOG08159	AT5G15400.1	U-box domain-c ontaining protein	AT5G15400.1	0	74,67	85,11	98	
PNG1 (Template: AT5G49570.1)	Glyma.03G246500.1.p	Peptide:N-glycanase	virNOG08136	AT5G49570.1	PNG1	AT5G49570.1	0	62,2	78,94	99	110
PNG1	Glyma.19G243900.1.p	Peptide:N-glycanase	virNOG08136	AT5G49570.1	PNG1	AT5G49570.1	0	57,33	72,41	87	110
OTU1*			virNOG00063	AT2G28120.1	OTU1						34
RAD23 (Template: AT1G16190)	Glyma.01G081500.1.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT3G02540.1	0	69,39	78,5	100	111

Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
RAD23	Glyma.07G002800.2.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT1G79650.4	0	67,68	77,78	100	111
RAD23	Glyma.08G223800.1.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT1G79650.4	0	65,78	77,18	100	111
RAD23	Glyma.08G262900.3.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT3G02540.1	0	68,53	77,86	100	111
RAD23	Glyma.10G275100.1.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT3G02540.1	0	69,39	78,27	99	111
RAD23	Glyma.13G371600.2.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT1G79650.4	7e-128	73,97	85,95	77	111
RAD23	Glyma.20G114700.3.p	Rad23 UV excision repair protein family	virNOG02197	AT1G16190.1 AT1G79650.4 AT3G02540.1 AT5G38470.1	RAD23A RAD23B RAD23C RAD23D	AT3G02540.1	1e-171	68,04	76,03	99	111
DSK2 (Template: AT2G17200.1)	Glyma.01G159500.1.p	Ubiquitin family protein	virNOG10597	AT2G17190.1 AT2G17200.1	DSK2A DSK2A	AT2G17200.1	0	57,95	72,08	100	89
DSK2	Glyma.02G041700.1.p	Ubiquitin family protein	virNOG10597	AT2G17190.1 AT2G17200.1	DSK2A DSK2A	AT2G17200.1	0	56,81	70,8	99	89
DSK2	Glyma.11G084700.1.p	Ubiquitin family protein	virNOG10597	AT2G17190.1 AT2G17200.1	DSK2A DSK2A	AT2G17200.1	0	57,77	71,55	100	89
DSK2	Glyma.16G118400.1.p	Ubiquitin family protein	virNOG10597	AT2G17190.1 AT2G17200.1	DSK2A DSK2A	AT2G17200.1	0	58,29	72,91	99	89
ER stress/plasma membrane cooperative response											
Upstream components											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
NAC062 (Template: AT3G49530.1)	Glyma.04G226700.1.p	GmNAC021 (Nomenclature Le et al. 2011)	virNOG05505	AT3G49530.1 AT4G35580.2 AT5G24590.2	ANAC062 CBNAC ANAC091	AT3G49530.1	2,00E-094	48,09	61,88	53	10
NAC062	Glyma.06G138100.1.p	GmNAC036 (Nomenclature Le et al. 2011)	virNOG05505	AT3G49530.1 AT4G35580.2 AT5G24590.2	ANAC062 CBNAC ANAC091	AT3G49530.1	4,00E-091	48,47	61,96	51	10

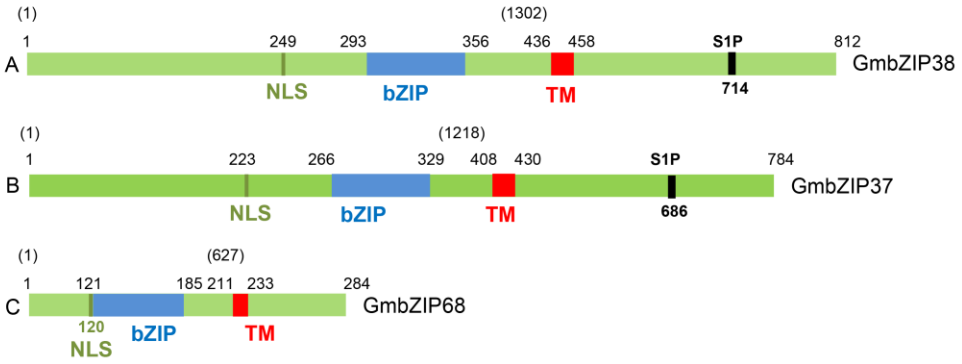
Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
NAC062	Glyma.14G189300.1.p	GmNAC110 (Nomenclature Le et al. 2011)	virNOG05505	AT3G49530.1 AT4G35580.2 AT5G24590.2	ANAC062 CBNAC ANAC091	AT3G49530.1	1,00E-096	48,49	61,75	52	10
Dowstream components											
Folging chaperones as in Table 1 - Molecular chaperones											



Additional file 2. Phylogenetic tree based on IRE-like sequences from Arabidopsis, soybean and rice. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.

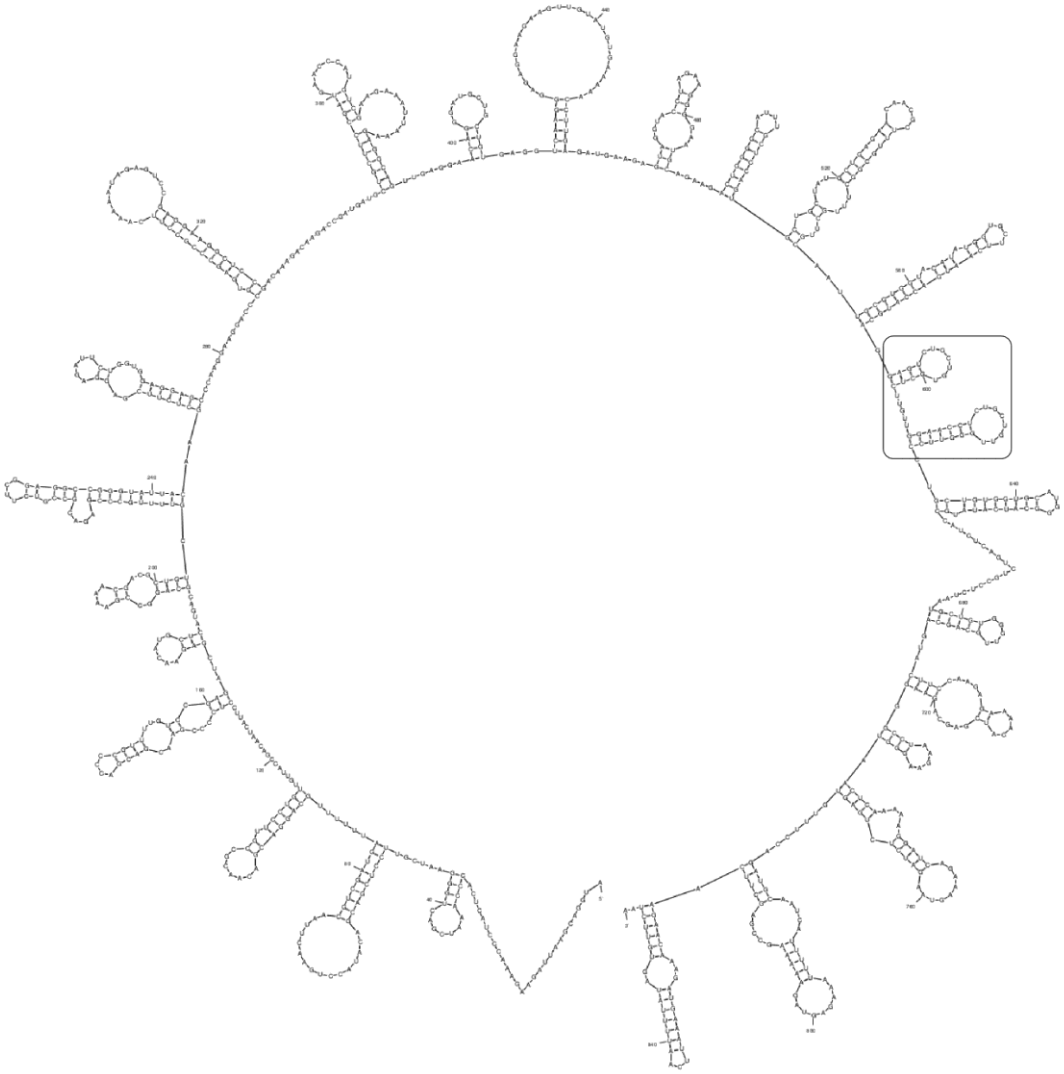


Additional file 3. Phylogenetic tree based on membrane-tethered bZIP-like sequences from Arabidopsis, soybean and rice. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model using MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.

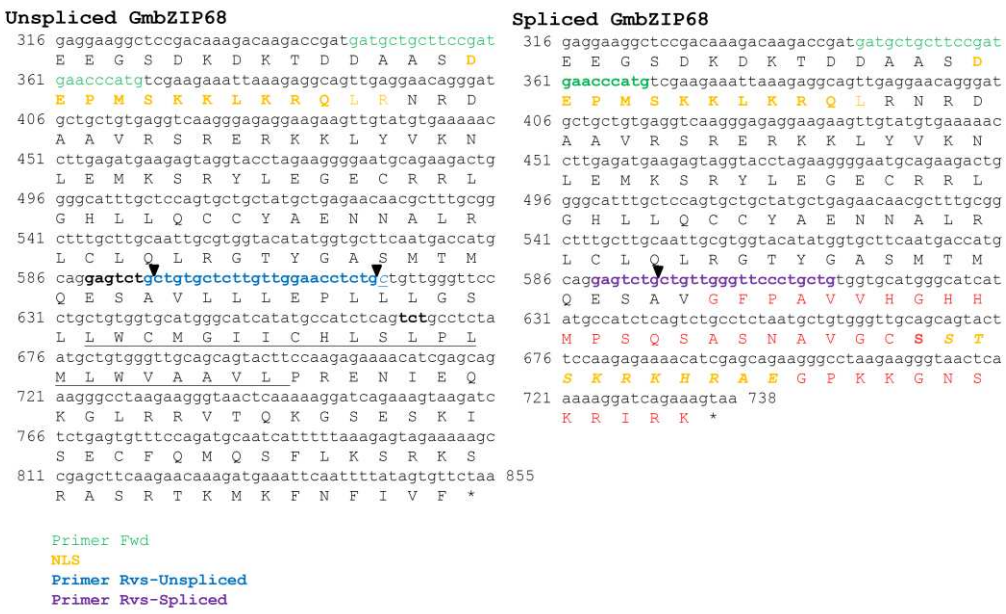


Additional file 4. Illustrative scheme of the predicted bZIP38 (A), bZIP37 (B) and bZIP68 (C) primary structures. The numbers above the figure indicate the amino acid positions in the predicted protein, and the numbers in parentheses indicate the corresponding nucleotide positions in the cDNA sequence. The bZIP domain is denoted in blue, TM is the putative transmembrane segment, S1P is the position of a canonical site for site-1 protease, and NLS indicates the position of a nuclear localization signal.

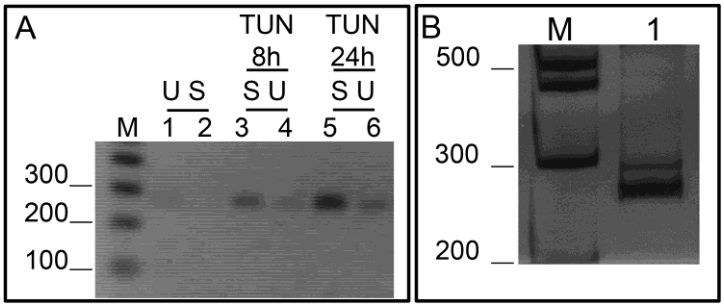
Additional file 5. Sequence alignments of bZIP17/28-like sequences from Arabidopsis and soybean. The sequence alignments of the indicated genes were obtained with CLUSTAL-W program. The bZIP domain, the transmembrane segment and a canonical S1P cleavage site are marked by open boxes.



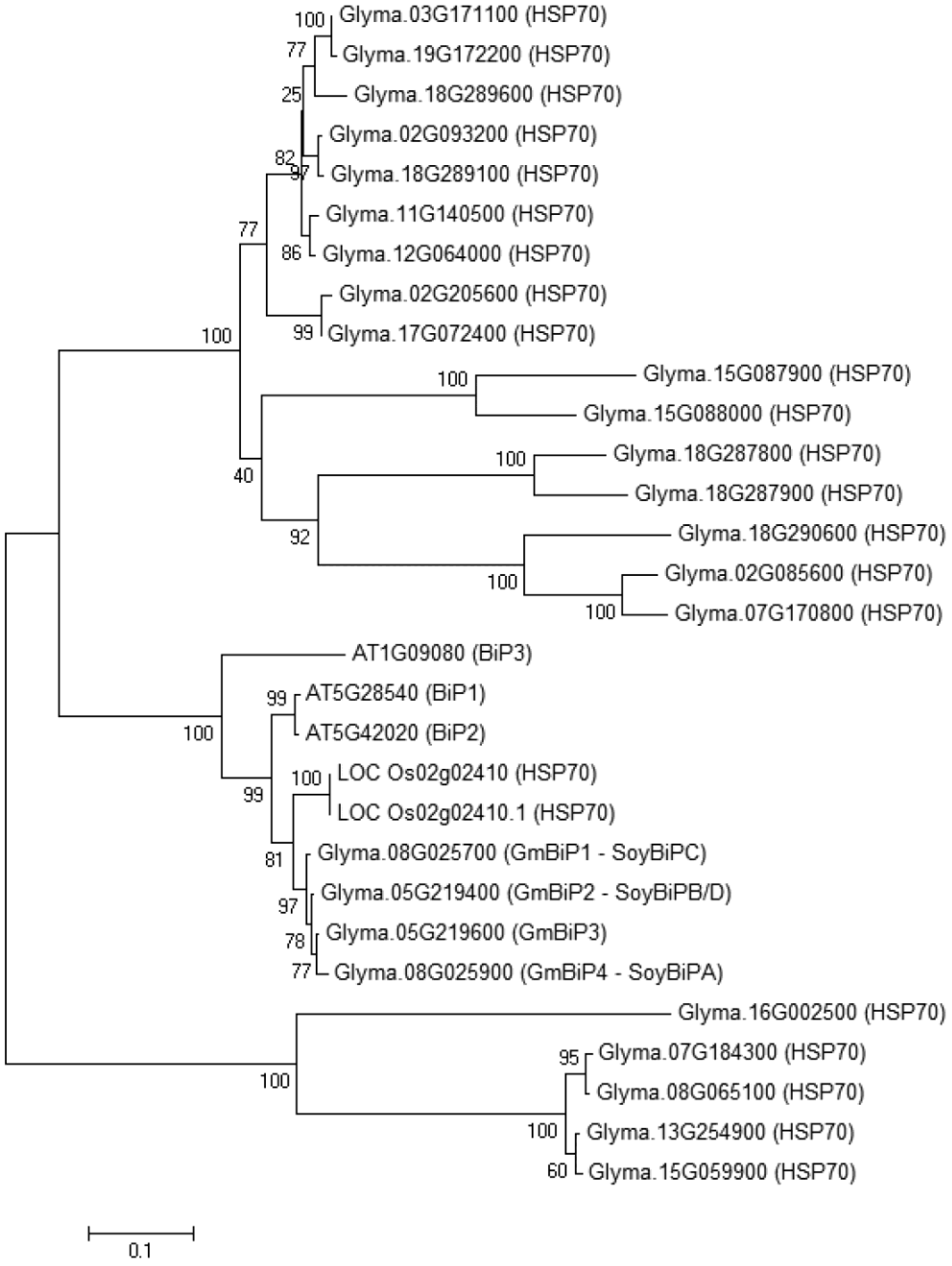
Additional file 6. Predicted structure of GmbZIP68 mRNA. The form of Glyma02g19754 mRNA folded by Mfold with the lowest free energy of ΔG = -191.80 (initially -187.80).



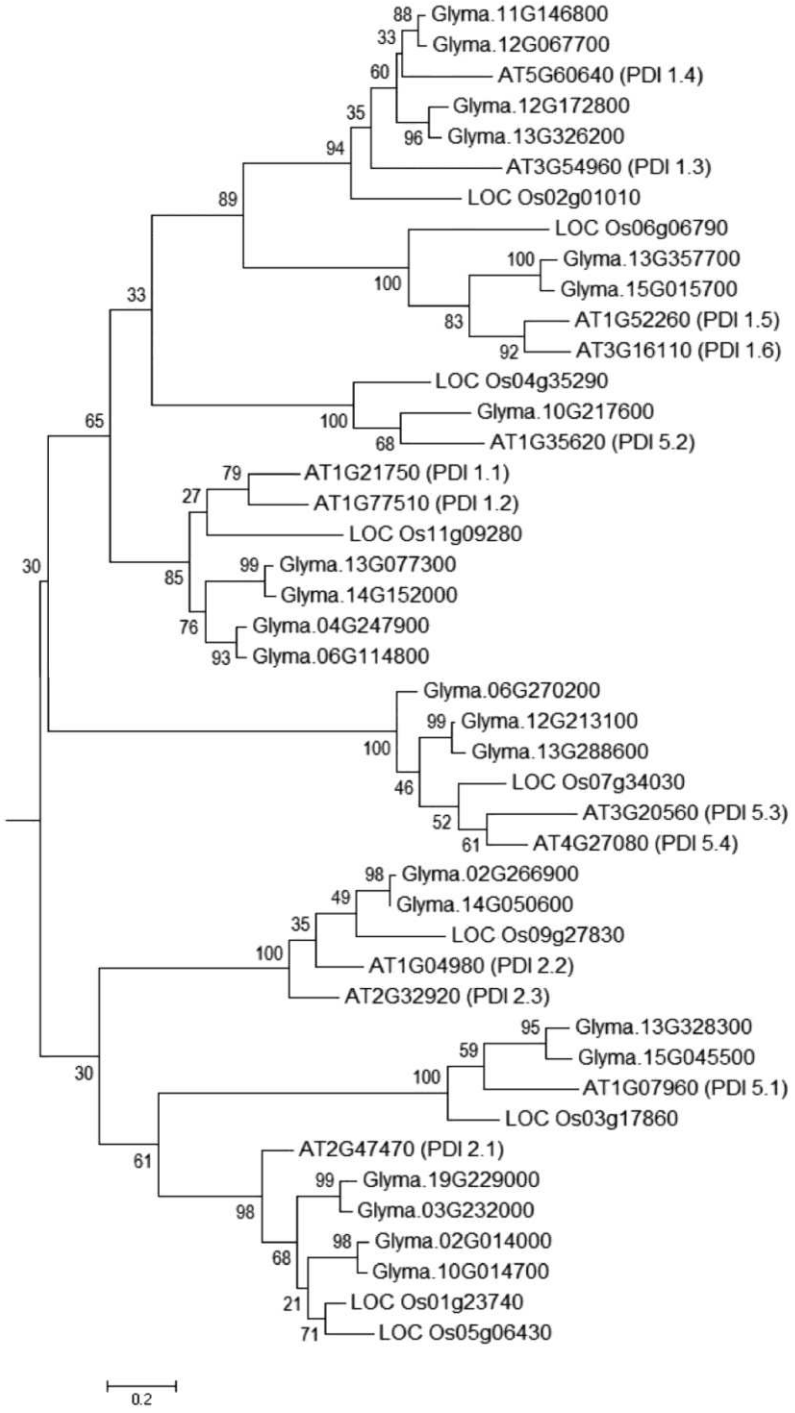
Additional file 7. Partial nucleotide and amino acid sequences derived from unspliced and spliced GmbZIP68 mRNAs. The arrows indicate the putative splicing sites in the unspliced mRNA and the ligation site in the spliced mRNA. The predicted nuclear localization signals (NLSs) are indicated by the amino acid sequences in orange. The predicted transmembrane segment is underlined. The amino acid sequence in red, derived from the spliced mRNA, shows the translational frameshift that resulted in a predicted amino acid sequence that was distinct from that of the unspliced mRNA. The nucleotide sequence in green corresponds to the forward primer, whereas the light blue sequence is complementary to the reverse unspliced primer, and the dark blue sequence is complementary to the reverse spliced primer used in splicing assay.



Additional file 8. Unconventional splicing of GmbZIP68 mRNA A. Electrophoretic patterns of RT-PCR products of ER stress-induced spliced GmbZIP68 mRNA on 1% agarose gels. Lanes 1 and 2 show the RT-PCR products generated using total RNA from untreated soybean seedlings with unspliced GmbZIP68 mRNA-specific primers (U, lane 1) and spliced mRNA-specific primers (S, lane 2). The RT-PCR products generated from RNA of soybean seedlings treated with tunicamycin for 8 h and 24 h are shown in lanes 3-6 using unspliced GmbZIP68 mRNA-specific primers (U, lanes 4 and 6) and spliced mRNA-specific primers (S, lanes 3 and 5). B. ER stress-induced unconventional splicing of AtbZIP60 mRNA. Total RNA from Arabidopsis seedlings treated for 6 h with tunicamycin was used as a template for RT-PCR performed with spliced bZIP60 mRNA-specific primers in combination with unspliced bZIP60 mRNA-specific primers.



Additional file 9. Phylogenetic tree based on HSP70-like sequences from *Glycine max* and *Arabidopsis thaliana*. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.



Additional file 10. Phylogenetic tree based on PDI-like sequences from *Glycine max* and *Arabidopsis thaliana*. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.

Additional file 11. A list of predicted/known orthologous genes involved in *Glycine max* and *Arabidopsis thaliana* plant-specific ER stress-mediated cell death pathways.

1) DCD/NRP-mediated cell death response											
Gene name (template Glycine max)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
GmERD15 (Template: Glyma02g42860.1)	Glyma.02G260800.1.p	No description	virNOG24368	Not have orthologous							38
GmERD15	Glyma.14G055200.1.p	No description	virNOG24368	Not have orthologous							
DCD/NRP-A (Template: Glyma20g16100.1)	Glyma.20G066100.1.p	DCD (Development and Cell Death) – NRP-A	virNOG03951	AT5G42050.1;	AtNRP1	AT5G42050.1	2,00E-112	52,27	64,8	96	37
DCD/NRP-A	Glyma.13G003200.1.p	DCD (Development and Cell Death) – NRP-A	virNOG03951	AT5G42050.1;	AtNRP1	AT5G42050.1	1,00E-107	52,82	64,88	96	
DCD/NRP-A	Glyma.08G024600.1.p	DCD (Development and Cell Death) – NRP-B	virNOG03951	AT5G42050.1;	AtNRP1	AT5G42050.1	6,00E-119	57,42	64,99	96	37
DCD/NRP-A	Glyma.05G218500.1.p	DCD (Development and Cell Death) – NRP-B	virNOG03951	AT5G42050.1;	AtNRP1	AT5G42050.1	1,00E-112	55,43	63,71	89	
DCD/NRP-C (Template: Glyma19G25870.1)	Glyma.16G056600.1.p	DCD (Development and Cell Death) – NRP-C	virNOG01663	AT3G27090.1;	AtNRP2	AT3G27090.1	8E-163	76	85	98	
DCD/NRP-C	Glyma.19G091100.1.p	DCD (Development and Cell Death) – NRP-C	virNOG01663	AT3G27090.1;	AtNRP2	AT3G27090.1	7E-163	75,67	84,33	98	
NAC081 (Template: Glyma12g02540.1)	Glyma.11G096600.1.p	GmNAC077 (Nomenclature Le et al.[97])	virNOG11218	AT2G17040.1;	anac036	AT2G17040.1	6E-120	59,25	73,29	96	39,54
NAC081	Glyma.12G022700.1.p	GmNAC081 (Nomenclature Le et al.[97])	virNOG11218	AT2G17040.1;	anac036	AT2G17040.1	2E-118	59,79	72,85	97	
NAC30 (Template: Glyma05g32850.1)	Glyma.05G195000.1.p	GmNAC030 (Nomenclature Le et al. [97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	1e-125	62,75	72,82	98	40
NAC30	Glyma.13G030900.1.p	GmNAC011 (Nomenclature Le et al. [97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	6e-146	70,88	78,6	97	40
NAC30	Glyma.04G208300.1.p	GmNAC018 (Nomenclature Le et al.[97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	2e-126	62,46	72,43	99	

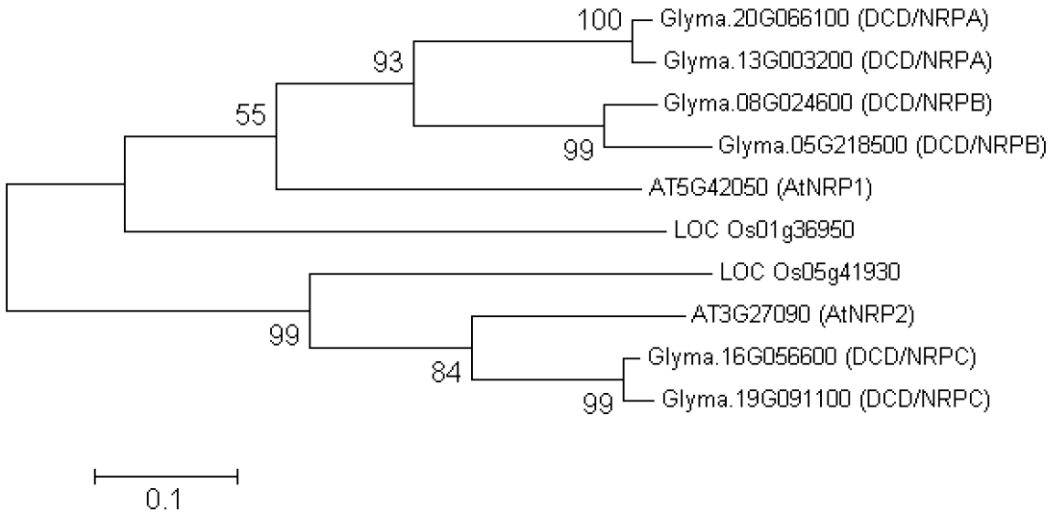
Gene name (template Glycine max)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
NAC30	Glyma.04G249000.1.p	GmNAC022 (Nomenclature Le et al. [97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	2e-148	69,97	79,21	98	
NAC30	Glyma.06G114000.1.p	GmNAC035 (Nomenclature Le et al. [97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	3e-150	69,54	79,8	98	
NAC30	Glyma.06G157400.1.p	GmNAC039 (Nomenclature Le et al. [97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	3e-125	62,25	71,19	99	
NAC30	Glyma.14G152700.1.p	GmNAC109 (Nomenclature Le et al. [97])	virNOG09836	AT1G01720.1;	ANAC002 ATAF1	AT1G01720.1;	1e-151	72,26	80,48	97	
VPE (Template: Glyma14g10620.1)	Glyma.14G092800.1.p	GAMMAVPE gamma vacuolar processing enzyme	virNOG04445	AT2G25940.1,] AT4G32940.1	ALPHA-VPE, GAMMA-VPE	AT4G32940.1;	0	77,09	86,78	94	40
VPE	Glyma.04G049900.1.p	GAMMAVPE gamma vacuolar processing enzyme	virNOG04445	AT2G25940.1, AT4G32940.1	ALPHA-VPE, GAMMA-VPE	AT4G32940.1;	0	76,67	87,9	95	
VPE	Glyma.06G050700.1.p	GAMMAVPE gamma vacuolar processing enzyme	virNOG04445	AT2G25940.1, AT4G32940.1	ALPHA-VPE, GAMMA-VPE	AT4G32940.1;	0	77,54	88,98	95	
VPE	Glyma.17G230700.1.p	GAMMAVPE gamma vacuolar processing enzyme	virNOG04445	AT2G25940.1, AT4G32940.1	ALPHA-VPE, GAMMA-VPE	AT4G32940.1;	0	76,09	85,43	95	

2) NAC89-mediated cell death signaling											
Upstream components											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
bZIP 17/28	Glyma.03G123200.1.p	bZIP transcription factor (bZIP38)	virNOG01396	AT2G40950.1; AT3G10800.1; AT3G56660.1;	bZIP17; bZIP28; bZIP49;	AT2G40950.1	8e-171	48,03	60,66	96	19, 20,26
bZIP 17/28	Glyma.19G126800.1.p	bZIP transcription factor (bZIP37)	virNOG01396	AT2G40950.1; AT3G10800.1; AT3G56660.1;	bZIP17; bZIP28; bZIP49;	AT2G40950.1	2e-167	47,68	61,32	94	19, 0, 26
IRE (Template: AT2G17520.1)	Glyma.01G157800.1.p	IRE1P-Related	virNOG09069	AT2G17520.1 ;AT5G24360.2;	ATIRE1a ;IRE1b	AT2G17520.1	0	68,86	82,27	70	12, 14
IRE	Glyma.09G197000.1.p	IRE1P-Related	virNOG09069	AT2G17520.1; AT5G24360.2;	ATIRE1a ;IRE1b	AT2G17520.1	0	66,59	81,94	49	

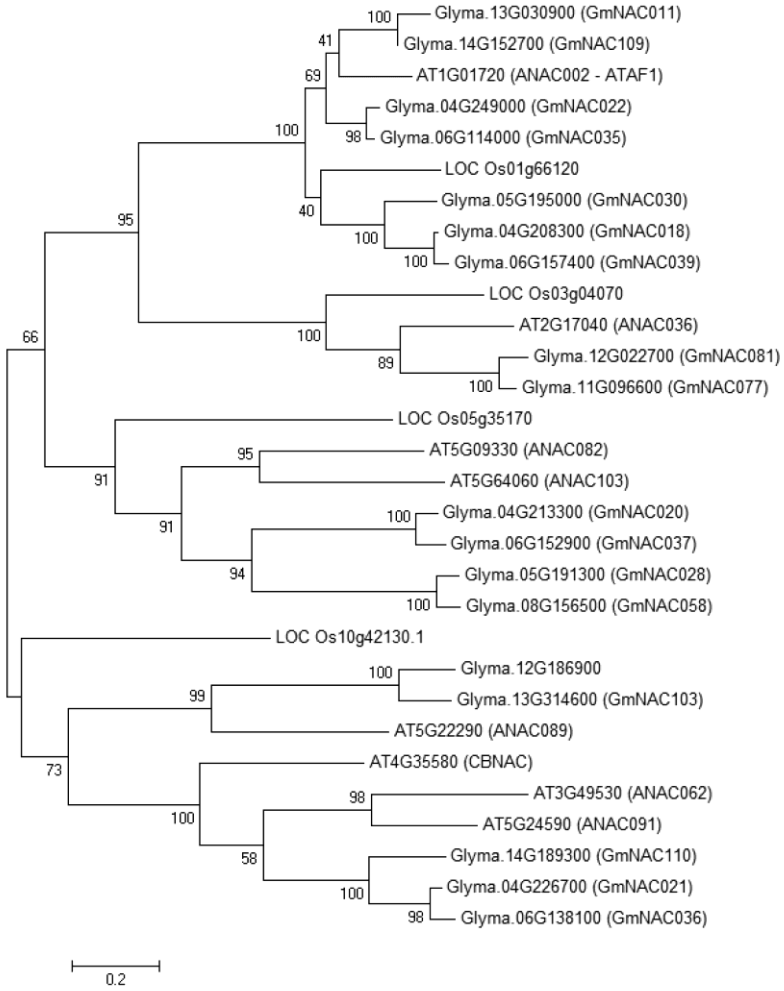
Gene name (template Glycine max)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Tair gene symbol	Best hit	E-value	Identity	Similarity	Coverage	Reference
IRE	Glyma.11G087200.1.p	IRE1P-Related	virNOG09069	AT2G17520.1; AT5G24360.2;	ATIRE1a; IRE1b	AT2G17520.1	0	65,73	80,17	74	
IRE	Glyma.16G111800.1.p	IRE1P-Related	virNOG09069	AT2G17520.1; AT5G24360.2;	ATIRE1a; IRE1b	AT5G24360.2	0	43,14	60,67	94	
Immediately Dowstream components											
Gene name (template Arabidopsis)	**Phytozome Glyma v.10.1.p**	**Description**	**EggNOG Group**	**eggNOG v4.0 *A. thaliana***	**Tair gene symbol**	**Best hit**	**E-value**	**Identity**	**Similarity**	**Coverage**	**Reference**
NAC89 (Template: AT5G22290.1)	Glyma.12G186900.1.p	unknown		AT5G22290.1;	ANAC89	AT5G22290.1	8,00E-089	51,14	72,35	52	36
NAC89	Glyma.13G314600.1.p	GmNAC103 (Nomenclature Le et al. [97])		AT5G22290.1;	ANAC89	AT5G22290.1	2,00E-086	48,35	67,03	72	
bZIP 60 (Template: AT1G42990.1)	Glyma.02G161100.1.p	GmbZIP68	euNOG19243	AT1G42990.1	ATBZIP60	AT1G42990.1	3,00E-020	46,62	57,43	52	14, 50

3) AGB1-mediated cell death signaling

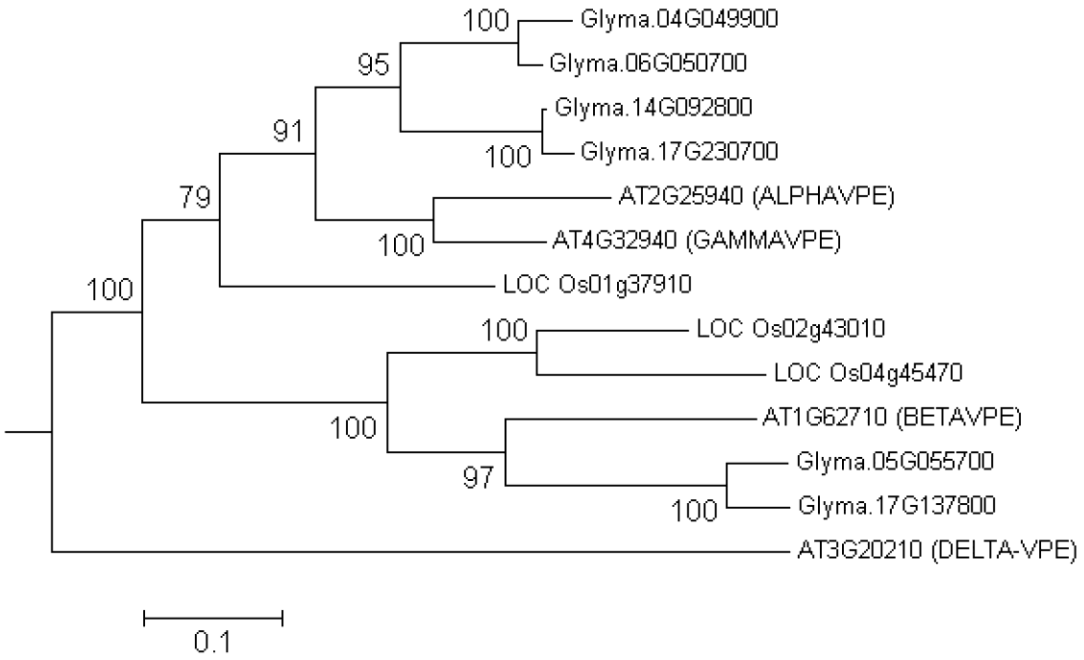
Gene name (template Arabidopsis)	Phytozome Glyma v.10.1.p	Description	EggNOG Group	eggNOG v4.0 *A. thaliana*	Best hit	Tair gene symbol	E-value	Identity	Similarity	Coverage	Reference
AGB1 (AT3G04480)	Glyma.04G013100.1.p	GTP binding protein beta 1	virNOG04773	AT4G34460	AT4G34460	AGB1	0,00E+000	80,37	90,45	99	112, 113
AGB1	Glyma.06G013000.1.p	GTP binding protein beta 1	virNOG04773	AT4G34460	AT4G34460	AGB1	0,00E+000	79,89	90,21	99	
AGB1	Glyma.11G118500.1.p	GTP binding protein beta 1	virNOG04773	AT4G34460	AT4G34460	AGB1	0,00E+000	81,38	90,16	99	
AGB1	Glyma.12G043900.1.p	GTP binding protein beta 1	virNOG04773	AT4G34460	AT4G34460	AGB1	0,00E+000	81,12	90,16	99	



Additional file 12. Phylogenetic tree based on DCD/NRP-like sequences from *Glycine max* and *Arabidopsis thaliana*. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.



Additional file 13. Phylogenetic tree based on NAC-like sequences from *Glycine max* and *Arabidopsis thaliana*. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.



Additional file 14. Phylogenetic tree based on VPE-like sequences from *Glycine max* and *Arabidopsis thaliana*. The unrooted phylogenetic tree was constructed using the maximum likelihood method with 10,000 bootstrap replications and the Jones-Talor-Thornton (JTT) amino acid substitution model with MEGA v.6 software. The numbers shown at the nodes indicate the percentage bootstrap scores.

Additional file 15. Primers used for PCR

Targeted sequence	Name	Primer sequence (5'- 3')
Glyma02g19754	Glyma02g19754GW-Fw	AAAAAGCAGGCTTCACAATGGACGAATTAGAAGAAAC
Glyma02g19754	Glyma02g19754GW-Rv	AGAAAGCTGGGTCACCCAACAGCAGAGGTTCCAAC
Glyma19g30681	Glyma19g30681GW-Fw	AAAAAGCAGGCTTCACAATGCTTAAAATAACTAACG
Glyma19g30681	Glyma19g30681GW-Rv	AGAAAGCTGGGTCCGTAGTTTTCCCCTCACTCTT
Glyma03g27865	Glyma03g27865Gw2-Fw	AAAAAGCAGGCTTCACAATGTACAGGGTTTTAGCGCG
Glyma03g27865	Glyma03g27865Gw2-Rv	AGAAAGCTGGGTCTTCTTA GTT TTC CCC TCA CTC TTC
attB1	2942	GGGGACAAGTTTGTACAAAAAAGCAGGCT
attB2	2943	GGGGACCACTTTGTACAAGAAAGCTGGGT
pDONR 207	3397 (Fwd)	TCGCGTTAACGCTAGCATGGATC
pDONR 207	3398 (Rvs)	TGTAACATCAGAGATTTTGAGACAC
35S	MC36 (Fwd)	TCCTTCGCAAGACCCTTCCTC
Glyma02g19754	glyma02g19754Fwd	GATGCTGCTTCCGATGAACCCATG
Glyma02g19754	glyma02g19754unspR	GCAGAGGTTCCAACAAGAGCACAG
Glyma02g19754	glyma02g19754spdR	CAGCAGGGAACCCAACAGCAGACTC

Additional file 16. List of gene-specific primers used for qRT-PCR

Primer	Sequence (5’ – 3’)	Gene
RT-Glyma19g30681FWD	GGTGCTGCTGGTGTCATGTG	Glyma19g30681
RT-Glyma19g30681-RVS	ATAGGTGCCATTGGAGGATGAT	Glyma19g30681
RT-Glyma03g27865-FWD	TCGACGACCTCGATGACATC	Glyma03g27865
RT-Glyma03g27865-RVS	GCGAAACGGAGGCGTAATTA	Glyma03g27865
RT-Glyma02g19754-FWD	GCTTCCGATGAACCCATGTC	Glyma02g19754
RT-Glyma02g19754-RVS	TTCCTCTCCCTTGACCTCACA	Glyma02g19754
HelicFw	TAACCCTAGCCCCTTCGCCT	HELIC
HelicRv	GCCTTGTCGTCTTCCTCCTCG	HELIC
BIPDFW	ATCTGGAGGAGCCCTAGGCGGTGG	BIPD
BIPDRV	CTTGAAGAAGCTTCGTCGTAAAACTAAG	BIPD
SMPFW	GCCGAACTGAGGAAAAGACGAACC	SMP
SMPRV	CTTGGGCTGTTTGTTGGGTCTTC	SMP
CALNFW	TGATGGGGAGGAGAAGAAAAAGGC	CNX
CALNRV	CACTTGGGTTTGGGATCTTGGCTC	CNX
Nac2 Fw	GGGTGCTTTGCCGTATTTACAA	GmNAC35
Nac2 Rv	CTCCTCCGCTTTTCAGAATCTC	GmNAC35
ActinFwd	AGCAAGGAAATTACCGCATTAGC	NbActin
ActinRvs	ACCTGCTGGAATGTGCTGAGA	NbActin