

BRUNO PAES DE MELO

**REVISITANDO A SUPERFAMÍLIA NAC NO GENOMA DA SOJA:
IDENTIFICAÇÃO E CARACTERIZAÇÃO DE NOVOS MEMBROS**

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada para a obtenção de título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2016

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

T

M528r
2016
Melo, Bruno Paes de, 1992-
Revisitando a superfamília NAC no genoma da soja :
identificação e caracterização de novos membros / Bruno Paes
de Melo. – Viçosa, MG, 2016.
vii, 63f. : il. (algumas color.) ; 29 cm.

Orientador: Elizabeth Pacheco Batista Fontes.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f.58-63.

1. Soja. 2. Genômica. 3. Genética vegetal. I. Universidade
Federal de Viçosa. Departamento de Bioquímica e Biologia
Molecular. Programa de Pós-graduação em Bioquímica
Aplicada. II. Título.

CDD 22. ed. 633.34

BRUNO PAES DE MELO

**REVISITANDO A SUPERFAMÍLIA NAC NO GENOMA DA SOJA:
IDENTIFICAÇÃO E CARACTERIZAÇÃO DE NOVOS MEMBROS**

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada para a obtenção de título de *Magister Scientiae*.

APROVADA: 26 de julho de 2016.

Paola Avelar Carpinetti-Oliveira

Juliana R. L S. Ramos
(Coorientadora)

Elizabeth P. B Fontes
(Orientadora)

AGRADECIMENTOS

Meus mais sinceros agradecimentos

a Deus, que sempre ouviu cada súplica minha, que sempre me guiou por caminhos retos e me ajudou em cada momento que eu pensei em fraquejar. A Ele, não toda honra e toda glória;

aos meus pais, Carmen e Duílio, e à minha irmã, Mariana. Minha família, suporte de todas as horas. Agradeço imensamente por sonharem comigo os meus sonhos e me auxiliarem na conquista deles, ainda que sacrifícios pessoais careçam de serem feitos;

à minha orientadora, Profa. Elizabeth Fontes, pelo companheirismo, confiança e pelos ensinamentos que fazem de mim um profissional melhor. Agradeço por sempre me estimular a alçar voos altos na ciência;

à minha coorientadora e amiga, Profa. Juliana Ramos (Ju), por me oferecer oportunidades que jamais me foram dadas e por sempre confiar em mim, acreditando no meu potencial enquanto bioquímico e pesquisador;

à minha noiva, Amanda, amor sempre presente, ombro amigo, por estar ao meu lado nos momentos difíceis e me lembrar que sou capaz. Por me auxiliar e me compreender;

aos amigos do LBMP, especialmente ao João Paulo e à Paola, pelo coleguismo, pelos conselhos, pelas críticas sempre construtivas e pelo suporte científico;

ao colega de trabalho e amigo, “Cleyninho”, pelo companheirismo nas análises de bioinformática, pelas dicas e pelo apoio;

aos técnicos do LBMP, Marlene, Gláucia e Adriano, pela presteza;

as agências de fomento, CNPq, CAPES e FAPEMIG, pelo suporte financeiro;

à Universidade Federal de Viçosa, ao BIOAGRO e ao INCTIPP pela estrutura disponibilizada para a execução deste trabalho.

SUMÁRIO

Resumo	vi
Abstract	vii
1. Introdução	01
2. Material e Métodos	07
2.1. Revisitando o genoma da soja em busca de proteínas com o domínio NAC	07
2.2. Construção da árvore filogenética dos genes da família NAC em soja	07
2.3. Distribuição dos genes da superfamília NAC nos cromossomos de soja	08
2.4. Ensaio de estresse em plântulas de soja.....	09
2.5. Extração de RNA total e síntese de cDNA	10
2.6. Clonagem dos genes da família NAC	11
2.7. Verificação da expressão transiente de NAC119 em <i>Nicotiana benthamiana</i>	13
2.8. Ensaio de transativação de NAC119 em sistema de duplo híbrido em leveduras	14

2.9. Quantificação da expressão dos novos genes da família NAC em condições de estresse abiótico e biótico -----	15
3. Resultados e Discussão -----	17
3.1. Análise <i>in silico</i> dos genes da superfamília NAC no genoma da soja ---	17
3.2. Clonagem de <i>GmNAC065</i> , <i>GmNAC085</i> e <i>GmNAC119</i> -----	28
3.3. Localização Celular de NAC119 -----	32
3.4. Avaliação da capacidade de transativação de NAC119 em sistema de duplo híbrido em leveduras -----	34
3.5. Análise do perfil da expressão gênica sob condições de estresse -----	34
4. Conclusão -----	43
5. Figuras Suplementares -----	45
6. Referências Bibliográficas -----	58

RESUMO

MELO, Bruno Paes de, M.Sc., Universidade Federal de Viçosa, Julho de 2016. **Revisitando a superfamília NAC no genoma da soja: identificação e caracterização de novos membros.** Orientadora: Elizabeth Pacheco Batista Fontes. Coorientadores: Juliana Rocha Lopes Soares Ramos e Pedro Augusto Braga dos Reis.

A recente liberação de uma nova anotação do genoma da soja (*Glycine max*) levantou a possibilidade da existência de novos genes pertencentes à superfamília NAC (*NAM*, *ATAF* e *CUC*). Os genes NAC codificam fatores de transcrição envolvidos no controle da morfofisiologia dos vegetais e nas respostas aos estresses ambientais. Neste trabalho, nós realizamos a revisitação do genoma da soja e o confrontamos com a estrutura altamente conservada do domínio NAC na busca de genes ainda não descritos na literatura. Nossa varredura identificou 32 novos genes candidatos, atualizando a superfamília para 180 membros. Também realizamos uma análise filogenética e agrupamos todos os genes em cinco grandes subfamílias, baseado na ortologia com genes de *Arabidopsis thaliana* e na proximidade filogenética com genes de soja cujas funções já foram caracterizadas. Por fim, procedeu-se à clonagem de três representantes da superfamília, sendo dois já descritos e um membro novo, que foi avaliado quanto à sua localização nuclear e sua capacidade de transativação em leveduras. Para validar as análises *in silico* da presença de novos genes NAC no genoma da soja e sua relação com respostas fisiológicas, quatro genes tiveram seus perfis de expressão gênica em condições de estresse monitorados por qRT-PCR, sendo que três deles eram novos membros encontrados nesta investigação. Estes resultados forneceram evidências de que os novos genes são provavelmente expressos e fundamentaram um inventário mais completo da superfamília de NAC da soja para futuras análises.

ABSTRACT

MELO, Bruno Paes de, M.Sc., Federal University of Viçosa, July, 2016. **Revisiting the NAC superfamily in soybean genome: identification and characterization of novel members.** Advisor: Elizabeth Pacheco Batista Fontes. Co-advisors: Juliana Rocha Lopes Soares Ramos and Pedro Augusto Braga dos Reis.

The release of a new soybean (*Glycine max*) genome-annotation raised the possibility that new NAC-superfamily genes would be present in soybean genome. The NAC (NAM, ATAF and CUC) genes encode transcription factors involved with the control of normal plant morphophysiology and stress responses. Here, we interrogated the newly annotated soybean genome against a conserved NAC-domain structure. Our data show 32 putative novel NAC-gene members, updating the superfamily to 180 gene members. We also organized the genes in five phylogenetic groups, according to the functions of orthologous *Arabidopsis thaliana* genes and NAC soybean genes. Finally, we cloned three gene members: two gene members already described and a new gene, which was characterized regarding to its nuclear localization and transactivation capacity. To validate our *in silico* analyses, we monitored the gene-expression profiles of three new NAC-genes by qRT-PCR. These data provided evidence that the new genes are probably expressed and established a more complete framework of the soybean NAC superfamily for future analyses.

1. Introdução

Sendo a economia brasileira sumariamente pautada na agroindústria, o estudo da fisiologia dos vegetais frente as diferentes condições de cultivo e estresses ambientais é de fundamental importância para a melhoria da produtividade agrícola e manejo de culturas. Sabe-se que os vegetais são organismos sésseis e estão sujeitos aos estresses do meio ambiente, como os estresses bióticos (causados por microrganismos, herbívoros ou vírus) e abióticos (disponibilidade de água, nutrientes, presença de agentes estressantes no solo, etc).

Nesse contexto, a adaptabilidade das monoculturas como a soja e o milho ganha destaque devido à grande área de plantio desses vegetais distribuída ao longo do território nacional e a importância econômica associada a elas. Segundo dados do governo brasileiro (disponíveis em Portal Brasil - www.brasil.gov.br/soja), a soja é a monocultura que gerou a maior contribuição nas taxas de exportação, chegando a casa dos R\$ 4 bilhões desde 2013. Estima-se que entre 2015 e 2016 o Brasil deva colher cerca de 198 milhões de toneladas, sendo o estado do Mato Grosso o maior produtor, contribuindo com aproximadamente 24% da safra nacional.

Tomando-se como justificativa a relevância econômica aliada à percepção das mudanças climáticas, que são fatores que exigem um manejo especial das culturas e a rápida aclimação das plantas, entender os mecanismos bioquímicos pelos quais os vegetais respondem ao estresse é crucial na intervenção biotecnológica que pode ser prospectada para obtenção de cultivares superiores.

Considerando alvos moleculares que interferem na resposta das plantas aos estresses, os fatores de transcrição se colocam em posição de destaque nos estudos da engenharia genética para a melhoria da agricultura. Associada a esta premissa, a principal propriedade dos fatores de transcrição é a capacidade de alterar ou reprogramar a expressão de genes envolvidos com a fisiologia e sinalização celular do vegetal, visando alterar o estado fisiológico da planta para uma maior adaptabilidade as condições ambientais.

Atualmente, mais de 5000 fatores de transcrição já foram identificados no genoma de soja [1], estando os fatores de transcrição da superfamília NAC (acrônimo para *NAM*, *ATAF* e *CUC*) entre os mais bem caracterizados. Constituída

de mais de 152 membros em soja [2], 151 membros em arroz (*Oryza sativa*) e 126 membros em *Arabidopsis thaliana* (<http://plntfdb.bio.uni-potsdam.de>), esses transfatores são caracterizados por possuírem um domínio amino-terminal conservado, responsável pela ligação ao elemento *cis* no DNA e um domínio carboxi-terminal variável em composição e tamanho, mas sabidamente responsável pela atividade reguladora da expressão gênica [3,4]. As proteínas NACs estão relacionadas com a regulação da morfofisiologia vegetal, atuando no controle dos processos de florescimento, embriogênese, sinalização hormonal e desenvolvimento de meristemas e respostas a estresses bióticos e abióticos [1-5]. Organizacionalmente, o domínio NAC contém cerca de 150-200 aminoácidos conservados, distribuídos em regiões denominadas subdomínios de A - E, sendo codificado por 2 éxons. Estudos detalhados de ligação entre NAC e o *cis* elemento regulador mostram que a sequência de reconhecimento de NAC (NACR) é do tipo “GACG” [3] (Figura 1).

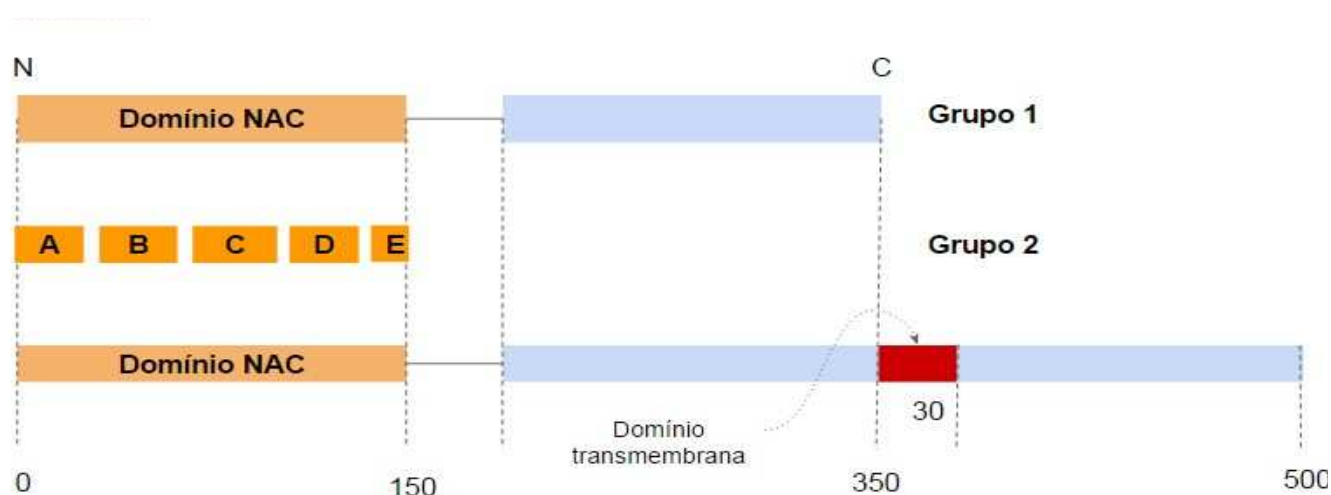


Figura 1. Diagrama esquemático representando a organização estrutural das proteínas da família NAC. A porção aminoterminal, responsável pela ligação ao elemento responsivo no DNA, é dividida em subdomínios de A-E. A porção carboxiterminal é variável em tamanho e pode conter sequências transmembrana. (Figura construída com base em Olsen *et al.* 2005).

Um dos primeiros genes da família de NAC a ser caracterizado foi o gene *NAM*, cujo nome fornece a primeira letra para o acrônimo NAC. Petúnias mutantes *nam* (*no apical meristem*), cujo meristema apical não se desenvolve, morrem muito

jovens ou, quando escapam, possuem um fenótipo aberrante relacionado ao desenvolvimento floral [7]. Em seguida, os genes *CUC1* e *CUC2* de *Arabidopsis* foram caracterizados devido à observação de mutantes que exibem sintomas semelhantes às petúnias *nam*. Além disso, nas plantas CUC deficientes, os cotilédones se apresentam fusionados, com forma de xícara (CUC - *cup-shaped cotyledon*) [8-9].

Quando expressos em mutantes nulos correspondentes, NAM, CUC1 e CUC2 fazem com que a planta recupere o fenótipo selvagem, o que demonstra o envolvimento desses genes na manutenção dos meristemas apicais e dos cotilédones. Todas essas proteínas têm em comum o domínio conservado NAC. Por isso, as proteínas CUC forneceram a terceira letra para o acrônimo NAC. Posteriormente, a expressão de um gene de *Arabidopsis* chamado *NAP* (*NAC-like activated by Apetala/Pistilatta*) mostrou que a proteína codificada continha um domínio semelhante ao NAC e era ativada por um heterodímero de fatores de transcrição relacionados à identidade do órgão floral [9].

Complementando os exemplos de funções biológicas essenciais ao desenvolvimento da planta, estudos mostraram que *NAC001* é superexpresso na ponta da raiz e nos pontos de desenvolvimento de raízes laterais. Bioquimicamente, esse gene é induzido pelo hormônio auxina, e o crescimento ou diminuição do número de raízes laterais varia de acordo com a expressão positiva ou negativa de *NAC001*, respectivamente. Genes que controlam o desenvolvimento de raízes, como o *AIR3* (*auxin-induced in root cultures*), possuem no promotor elementos de reconhecimento para *NAC001* [10].

Os fatores de transcrição que respondem aos estresses, dentre os quais se encontram os DREBs (*dehydration responsive element binding proteins*), os AREBs (*ABA responsive element binding proteins*) e os NACs, atuam em cascatas coordenadas para montar uma resposta de defesa. Correlações têm sido reportadas entre a expressão de NACs e de genes *ERDs* (*early responsive to dehydration*), que são transfatores responsivos precoces na sinalização do estresse hídrico [11-16]. Esses genes são induzidos por ABA, ácido metil-jasmônico e outros fitormônios sinalizadores da seca e da alta salinidade e, portanto, devem participar em mecanismos de defesa a estresses abióticos [15,16].

Em soja, ainda não existe uma robustez dos dados de caracterização estrutural e fisiológica do grande número de NACs em seu genoma. Baseado na homologia de alguns genes NAC de soja com genes de *Arabidopsis*, tabaco, batata, arroz e outras espécies, têm-se feito predições das funções biológicas dos NACs de soja e os colocados em grupos de acordo com a função de seus ortólogos.

Os estresses abióticos promovem uma série de respostas nas plantas desencadeando mecanismos que adaptam a fisiologia do vegetal à adversidade para garantir sua manutenção. Para isso, vias metabólicas são ativadas pela reprogramação da expressão gênica, coordenada por múltiplos sinais que envolvem fitormônios, proteínas cinases, fosfatases e fatores de transcrição. Em diversas espécies de planta já se tem demonstrado o papel das proteínas NAC na coordenação dessas respostas. Linhagens transgênicas de arroz superexpressando *OsNAC2/6* e *OsNAC10* mostram aumento na tolerância ao estresse salino e a seca [17-18]. A superexpressão de *OsNAC10*, *OsNAC5* e *SNAC1* levou ao aumento do diâmetro de vasos na raiz, permitindo a maior absorção de água com consequente aumento de tolerância à seca e maior rendimento de grãos sob estresse hídrico [19-20]. Recentemente, o gene *SNAC3*, já associado ao aumento da tolerância ao calor e à seca, mostrou ser um regulador da tolerância aos estresses oxidativos, uma vez que é capaz de regular a expressão de genes associados à geração de espécies reativas de oxigênio (EROS) [21]. Em *Arabidopsis*, perfis de expressão gênica em diferentes condições de estresse e em diferentes estágios de vida da planta demonstram a plasticidade dos transfatores NAC no controle da expressão gênica em diferentes processos e estresses fisiológicos. Os fatores de transcrição NTL4 e JUB-1, ambos do tipo NAC, têm envolvimento com as vias de sinalização mediadas por EROS. NTL4 é capaz de se ligar diretamente aos *cis* elementos na região promotora de genes *Atrboh* levando ao aumento da concentração de EROS e a senescência foliar, enquanto JUB-1 é responsivo ao peróxido de hidrogênio e pode se ligar a promotores de genes *DREB* (*dehydration responsive element binding protein*) [22-23]. A superexpressão de outros fatores, como ANAC019, ANAC055 e RD25 (ANAC072) conferiu maior tolerância ao estresse hídrico [15, 24].

Em outras culturas como o pinheiro bravo (*Pinus pinaster*) e soja, os genes da família NAC executam papéis semelhantes àqueles descritos em arroz e *Arabidopsis*. Em pinheiro, 37 fatores de transcrição contendo o domínio NAC foram

identificados e os genes *PpNAC3* e *PpNAC2* mostraram-se responsivos aos estresses por ferimento, baixa temperatura e salinidade, sendo controlados por fatores de transcrição do tipo MYC responsivos ao jasmonato, também elicitor das respostas aos estresses bióticos [25]. Dados de análise de microarranjo de DNA em plantas de soja sob condições de déficit hídrico mostraram que diversos genes da família NAC são diferencialmente expressos em resposta a diminuição do potencial hídrico nas folhas, incluindo *GmNAC085*, *GmNAC019* e *GmNAC092* [2, 26-27].

Dentre as diversas respostas elicitadas na planta sob condições de estresse, inclui-se a senescência foliar. Esse processo é caracterizado pela transição do tecido foliar com o aparelho fotossintético ativo para uma extensiva degradação tecidual, com conseqüente amarelecimento seguido de necrose. Esse fenômeno é controlado por genes associados à senescência (SAG – em inglês, *senescence-associated genes*), nos quais se incluem alguns genes da família NAC [3, 28]. Em *Arabidopsis*, a família SNAC-A, compreendida por 7 genes nos quais se incluem *ATAF1/2*, está envolvida com a senescência foliar em resposta coordenada com o hormônio ABA e, frequentemente, tem sido associada aos estresses bióticos. A superexpressão de *ATAF-2* resulta na maior suscetibilidade ao fungo necrotrófico *Fusarium oxysporum* devido à repressão de genes PR, mas também aumenta a resistência à penetração de fungos biotróficos, como *Blumeria graminis* [29-30]. Recentemente, o mecanismo de regulação da senescência em *Arabidopsis* mediado por ANAC092/ORE1 foi descrito. A via possui um braço de regulação temporal mediado por um microRNA: envelhecimento da folha ou condições de estresse levam a expressão da proteína EIN2 (*ethylene insensitive 2*), que promove o decaimento do miRNA164, cujo alvo para a clivagem é ANAC092/ORE1. Quando expresso, ele ativa genes relacionados à senescência e morte celular programada, além de funcionar como repressor de genes do tipo GK, fundamentais à manutenção dos cloroplastos [31].

Pesquisas que visam compreender como os fatores de transcrição da grande família dos NACs atuam integrando as vias de sinalização para estresse osmótico e de retículo endoplasmático, identificaram genes que convergem em nível transcricional para a ativação da via de morte celular programada (PCD, *programmed cell death*) em soja [5, 11-14, 32-33]. Os componentes dessa cascata de sinais incluem o transfator GmERD15, as proteínas DCD/NRP (*development and*

cell death domain-containing N-rich proteins) e os fatores do tipo NAC, GmNAC81 e GmNAC30, que interagem entre si para ativar a expressão da *caspase-like* VPE (*vacuolar processing enzyme*), que executa o processo de morte celular programada específico de plantas por meio do colapso do vacúolo, com consequente fragmentação do DNA, perda de clorofila, degradação de proteínas e aumento da peroxidação de lipídeos [5, 11, 32-33]. Foi observado que *GmNAC081* também tem envolvimento com o processo de senescência natural e seu silenciamento retarda a senescência foliar e atenua os demais sintomas decorrentes da morte celular programada [34].

Uma relação de convergência entre o estresse osmótico e no retículo endoplasmático na soja também revela o envolvimento de transfatores NAC na regulação dessas respostas. O chaperone molecular BiP, residente no retículo endoplasmático, funciona como modulador negativo da via de sinalização de morte celular mediada por NRPs e, como consequência, confere tolerância à seca quando superexpresso em plantas transgênicas. BiP é capaz de modular a atividade de GmNAC30 e GmNAC81 no circuito GmNAC81/30/VPE [5, 11, 32-33].

Recentemente, uma nova anotação do genoma da soja foi disponibilizada em um banco de dados de genomas vegetais (Phytozome[®] - www.phytozome.jgi.doe.gov), permitindo a revisitação do genoma da soja e o remapeamento de todos os genes da superfamília NAC. Fundamentado na hipótese de que o entendimento e intervenção na expressão de fatores envolvidos na resposta vegetal ao estresse podem levar ao desenvolvimento de estratégias biotecnológicas que garantam maior tolerância às adversidades, o presente trabalho reidentificou e remapeou todos os genes da família NAC em soja, agrupando-os filogeneticamente, além de correlacionar alguns membros com as respostas aos estresses biótico e abiótico.

2. Material e Métodos

2.1. Revisitando o genoma da soja em busca de proteínas com o domínio NAC

O *Phytozome*® é um portal comparativo de genomas de plantas desenvolvido pelo Departamento de Energia do *Joint Genome Institute* - Universidade da Califórnia, Estados Unidos - com 58 genomas de plantas sequenciados e anotados, dentre os quais se encontra o genoma da soja (About Phytozome - in www.phytozome.jgi.doe.gov).

A recente liberação de uma nova versão de montagem do genoma da soja levantou a possibilidade de que novos genes da superfamília NAC com ORFs completas estivessem disponíveis. Para a análise de possíveis novas proteínas, o domínio NAC, já bem caracterizado e descrito no Pfam (29.0 - <http://pfam.xfam.org/>), foi utilizado como protótipo para a varredura contra o genoma completo da soja na versão Wm82.a2.v1 (montagem disponível na versão v11.0 do *Phytozome* - janeiro de 2016), anotado no formato FASTA. A comparação das sequências foi feita no software InterProScan 5 (<http://www.ebi.ac.uk/Tools/pfa/iprscan5/>) e as novas proteínas encontradas, não descritas pela varredura mais completa da superfamília NAC de soja publicada por Le *et al.* (2011), foram agrupadas em uma tabela contendo o *ID number* (nomenclatura padrão: Glyma v11.0), o tamanho da proteína, a posição dos aminoácidos de início e fim do domínio NAC, o E-value do contraste e a presença de domínio transmembrana com sua respectiva probabilidade.

2.2. Construção da árvore filogenética dos genes da família NAC em soja

Para a construção da árvore filogenética dos genes NAC de soja, as sequências dos genes da mesma família em *Arabidopsis thaliana* foram obtidas no *Phytozome*® e utilizadas no alinhamento global com todas as sequências de soja

previamente obtidas utilizando o software *Muscle* (<http://www.ebi.ac.uk/Tools/msa/muscle/>).

O alinhamento múltiplo entre as sequências dos genes de *Arabidopsis* e de soja permitem uma comparação não apenas baseada na conservação de sequências, mas também baseada em respostas fisiológicas desencadeadas por esses genes, já que as proteínas com domínio NAC em *Arabidopsis* são bem caracterizadas na literatura. Para refinar o agrupamento, o banco de proteínas ortólogas *EggNOG* 4.5 (<http://eggnogdb.embl.de/#!/app/home>) foi utilizado para a comparação de cada proteína de soja com cada proteína de *Arabidopsis*.

Utilizando o software *Fasttree* (<http://meta.microbesonline.org/fasttree/>), que tem como método estatístico a verossimilhança em um consenso de 10.000 *bootstrap*, procedeu-se a reconstrução da árvore, visualizada e editada no programa *FigTree* (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.3. Distribuição dos genes da superfamília NAC nos cromossomos de soja

Para avaliar a organização dos genes da família NAC nos cromossomos da espécie *Glycine max*, fez-se o mapeamento dos genes baseado na sequência do domínio NAC disponível sob o código IPR003441 no banco de dados do InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence-search>). A análise foi feita com auxílio do software *PLAZA* 3.0 (Plant Comparative Genome - http://bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots/genome_mapping/index), fornecendo um diagrama esquemático dos cromossomos de soja e das posições de todos os genes da família NAC, que, de acordo com os códigos ID dos genes no Phytozome ®, foram manualmente nomeados utilizando o programa *CorelDraw* ® (COREL).

2.4. Ensaio de estresse de plântulas de soja

Uma vez que os fatores de transcrição são induzidos frente a situações de estresse, a estratégia para a amplificação e isolamento de novos genes da família NAC de soja consistiu em estressar plântulas em estágio de desenvolvimento entre V3 e V4, crescidas em casa de vegetação (12 horas de fotoperíodo, temperaturas de 15 °C - 30 °C, 70% de umidade relativa) com agentes que simulam os principais estresses ambientais, sendo o estresse osmótico, o de retículo endoplasmático e o estresse biótico induzidos pelo tratamento das plântulas com os compostos, concentrações e tempos descritos na **Tabela 1**.

Tabela 1. Tratamentos utilizados no ensaio de estresse de plântulas de soja. Para induzir respostas fisiológicas semelhantes aos estresses de seca, retículo endoplasmático e estresses bióticos foram administrados os compostos PEG, Tunicamicina e Ácido Salicílico, respectivamente.

Tratamento	Efeito de Estresse Induzido	Tempo de Tratamento	Concentração
Polietilenoglicol (PEG)	Osmótico	0h; 0,5h; 2h e 12h	0,5% (v/v)
Tunicamicina	Retículo Endoplasmático	0h; 0,5h; 2h e 12h	5 ug/mL
Ácido Salicílico	Biótico	0h; 0,5h; 2h e 12h	5 mM

As soluções para o ensaio foram preparadas utilizando a solução nutritiva de Hoagland como solvente para minimizar as respostas de estresse induzidas pela retirada da planta do solo. A solução de tunicamicina foi preparada utilizando uma solução estoque do composto diluído em DMSO (dimetilsulfóxido) na concentração de 50 ug/mL e a solução de ácido salicílico foi preparada dissolvendo-se 0,497g do composto em 500 uL de DMSO e completando o volume com solução de Hoagland para 720 mL. No caso da solução de PEG, utilizou-se uma solução estoque de 10% (m/v) do composto diluído em água. Para os controles (tempo de 0h), utilizou-se apenas solução nutritiva de Hoagland acrescida de DMSO (para os tratamentos com

tunicamicina e ácido salicílico) ou a solução nutritiva pura (para o tratamento com PEG).

Em cada ensaio, foram utilizadas 12 plantas no mesmo estágio de desenvolvimento e com as mesmas características (tamanho do caule, tamanho das folhas e tamanho da raiz), aclimatadas por 24 horas em frascos individuais com 15 mL de solução nutritiva de Hoagland.

Após o tempo de tratamento, as plantas foram retiradas da solução e discos foliares com diâmetro de 1 cm foram coletados, armazenados em tubos eppendorf e imediatamente congelados em nitrogênio líquido. O restante da parte aérea da plântula também foi coletado e congelado. Todo o material biológico foi armazenado a -80 °C em ultrafreezer após o ensaio.

2.5. Extração de RNA total e síntese de cDNA

Os discos foliares coletados no ensaio de estresse foram agrupados em 4 pools com 3 repetições biológicas em cada pool, dos quais foram extraídos o RNA total pelo método do Trizol[®] (Invitrogen). As amostras foram quantificadas em espectrofotômetro NanoDrop (Thermo Scientific) e 4 ug de RNA total em um volume máximo de 8 uL foram tratados com DNase I (Qiagen[®]) por 15 minutos (temperatura ambiente) em uma reação contendo o volume de RNA, 1 uL de tampão da enzima 10X e 1,5 uL de enzima (3 U). A inativação da enzima se deu com a adição de 1 uL de EDTA 25 mM. Após o tratamento, as amostras foram submetidas à eletroforese em gel de agarose para a verificação de sua integridade e procedeu-se à síntese do cDNA. O volume de 10 uL do RNA total foi utilizado para cada reação, que se procedeu em termociclador em duas etapas: 1. adição de uma solução composta de 1 uL de oligoDT (18T) (10 uM) e 1uL de dNTPs (10mM) seguido de incubação a 65 °C por 5 minutos e incubação a 4 °C por 5 minutos; 2. adição de uma solução composta 4 uL de tampão da enzima 5X, 2 uL de DTT (0,1 M), 1 uL de água RNase free (tratada com DEPC) e 1 uL de enzima transcriptase reversa MMLV (Invitrogen) (200 U), seguindo de incubação a 37 °C por uma hora, 70 °C por 15 minutos e 4°C até a retirada das amostras.

2.6. Clonagem dos genes da família NAC

Para a clonagem dos genes da família NAC, as sequências das ORFs foram obtidas do *Phytozome*® e utilizadas para o desenho e síntese de oligonucleotídeos iniciadores adaptados ao sistema *Gateway*® *cloning technology* (Invitrogen) descritos na **Tabela 2**.

Tabela 2. Oligonucleotídeos iniciadores utilizados na clonagem dos genes NAC pelo sistema *Gateway* (Invitrogen).

Oligonucleotídeo	Sequência (5' - 3')	Código do Gene
Glyma NAC 85 fwd	AAAAAGCAGGCTTCACAA TGGGAAGTTCAGAGAGAGA	Glyma.12G149100
Glyma NAC 85 rev st	AGAAAGCTGGGTCTCAGT CCCCTAAACCCGAACTC	Glyma.12G149100
Glyma NAC 65 fwd	AAAAAGCAGGCTTCACAA TGGAGAAGGTGAAATTTTGTG	Glyma.08G360200
Glyma NAC 65 rev ns	AGAAAGCTGGGTCTTCTA ATAGTATAAGGGAAGG	Glyma.08G360200
NAC119 fwd	AAAAGCAGGCTTCACAAT GGAGAACATAATTTCCATATGTGGATC	Glyma.16G016400
NAC119 rev ns	AGAAAGCTGGGTCTGGGTA TGAACATCTTCGTCATCCGTAAC	Glyma.16G016400
attB1	GGGGACAAGTTTG TACAAAAAGCAGGCT	-
attB2	GGGGACCACTTTGTA CAAGAAAGCTGGGT	-

Procederam-se duas reações sequenciais de PCR (27 ciclos - volume final de 25 uL e 50 uL, respectivamente) contendo 1uL de cDNA (~ 200 ng), 1 uL de oligonucleotídeos fwd (10 uM), 1 uL de oligonucleotídeos rvs (10 uM), 4 uL de dNTPs (2,5 mM), 5 uL de tampão da enzima 10X, 4 uL de MgSO₄ (50 mM) e 0,2 uL de enzima *Platinum Taq High Fidelity polymerase* (Invitrogen) (2 U). As reações sequenciais consistiram na amplificação da ORF utilizando oligonucleotídeos específicos na primeira etapa e inserção das regiões de recombinação sítio dirigida

attB1 e attB2 na segunda etapa. Os oligonucleotídeos iniciadores utilizados na amplificação de cada ORF encontram-se descritos na Tabela 2.

Depois de amplificados, os insertos foram purificados diretamente do gel de agarose utilizando o kit *QIAquick Gel extraction Kit* (QIAGEN) e colocados em vetores de entrada pDONR 201[®] e/ou pDONR 207[®] (Invitrogen) por meio de recombinação sítio dirigida utilizando a enzima BpClonase (Invitrogen). Para cada reação, foram utilizados 50 fmol do inserto isolado, 50 fmol do vetor vazio e 1 uL de enzima. As reações foram incubadas a 25 °C por 8 horas e interrompidas com 1 uL de proteinase K (2 ug) por 10 minutos a 37 °C.

Os produtos da reação de recombinação foram transformados em células de *E. coli* linhagem DH5α eletrocompetentes e os transformantes selecionados em meio LB-ágar suplementado com canamicina (para o vetor pDONR201) ou gentamicina (para o vetor pDONR207), nas concentrações de 100 uM e 10 uM, respectivamente. Os DNAs plasmidiais das colônias positivas foram extraídos utilizando o kit *QIAprep Spin Miniprep kit* (QIAGEN) e enviados para sequenciamento automatizado de Sanger para a confirmação da identidade dos clones.

Paralelamente, por meio da reação de recombinação catalisada pela enzima LRClonase (Invitrogen), os insertos foram transferidos para o vetor de destino pDEST32[®] (*Two Hybrid Yeast System* - Invitrogen) e o vetor binário de expressão pK7FWG2 (Invitrogen). Os clones gerados encontram-se descritos na **Tabela 3**.

Tabela 3. Clones em *Escherichia coli* gerados com a amplificação dos genes NAC. Os clones foram gerados pela amplificação dos genes e recombinação sítio dirigida utilizando o sistema Gateway®.

Identificação do Clone	Descrição
pUFV 2589	NAC 65 ns* clonado em pDONR 201 a partir de cDNA de folhas de soja estressadas com tunicamicina (5 ug/mL) por 48 h.
pUFV 2826	NAC 119 ns* clonado em pDONR 201 a partir de uma mistura de cDNAs de plântulas de soja estressadas com PEG, Tunicamicina e Ácido Salicílico.
pUFV 2827	NAC 85 st** clonado em pDONR 207 a partir de uma mistura de cDNAs de plântulas de soja estressadas com PEG, Tunicamicina e Ácido Salicílico.
pUFV 2828	NAC 119 ns* clonado em pDEST 32 a partir de reação de recombinação via LR clonase utilizando o clone pUFV 2826.
pUFV 2830	NAC 119 ns* clonado em pK7FWG2 a partir de reação de recombinação via LR clonase utilizando o clone 2826. O clone resultante contém o cDNA de NAC 119 fusionado ao aminoterminal de GFP.

(*) ns - sem o códon de terminação

(**) st - com o códon de terminação

2.7. Verificação da expressão transiente de NAC119 em *Nicotiana benthamiana*

A construção de DNA pUFV2830, contendo a ORF clonada do gene *NAC119* (Glyma.16G016400) no vetor binário de expressão pK7FWG2, foi utilizado para a transformação de células de *Agrobacterium tumefaciens* linhagem GV3101 eletrocompetentes. Os transformantes foram duplamente selecionados em meio sólido LB-ágar suplementado com gentamicina (50 mg/mL) e espectomicina (100 mg/mL) para a preparação de uma suspensão de infiltração, contendo meio MES 100 mM (pH = 5,6), MgCl₂ 100 mM e acetoseringona 200 mM. As células foram

centrifugadas por 4 minutos a 4.500 rpm (Centrífuga Eppendorf 5424 R) e lavadas por 3 vezes com 2 mL da solução de infiltração.

A suspensão foi mecanicamente inoculada em folhas de *Nicotiana benthamiana* para a indução da expressão transiente da proteína NAC119 fusionada à extremidade aminoterminal da proteína GFP (*green fluorescent protein*). Após 3 dias da inoculação, quadros com 1 cm² foram excisados das folhas de tabaco, depositados em lâminas de vidro para microscopia, cobertos com água deionizada e observados ao microscópio confocal de varredura laser Zeiss LSM510 META sob feixe de laser nos comprimentos de onda de 498 nm para a excitação da proteína fusionada à GFP.

2.8. Ensaio de Transativação de NAC119 pelo sistema de duplo híbrido em leveduras

Para o ensaio de transativação utilizando o sistema de duplo híbrido em leveduras, foi utilizado o clone pUFV2828, contendo o gene NAC119 clonado em pDEST32. Esse clone permite a expressão do cDNA de NAC119 fusionado ao domínio de ligação BD (*bind DNA domain*) da proteína GAL4 em leveduras. Em leveduras com o gene *HIS3* sob o controle de um promotor contendo o sítio de ligação de GAL4, a capacidade de transativação da proteína analisada é monitorada pela auxotrofia em histidina.

A levedura *S. cerevisiae* linhagem AH 109 foi transformada com o clone pUFV2828 pelo protocolo PEG-acetato de lítio e os transformantes positivos foram selecionados em meio sólido SD deficiente em leucina (SD/-leu/), já que o vetor possui marca auxotrófica para esse aminoácido.

Para verificar a capacidade de transativação de NAC119, as colônias positivas foram crescidas por 8 horas em meio líquido SD/-leu/ até atingirem a densidade ótica (D.O.) superior a 1,0 e a cultura foi diluída em meio SD/-leu/ até D.O. igual a 1,0, 0,1 e 0,01. Todas as diluições foram plaqueadas em gotas de 10 uL em meio sólido SD deficiente em leucina e histidina, acrescidas de 3-aminotriazol nas concentrações de 0 mM, 2,5 mM, 5,0 mM, 7,5 mM e 10 mM para verificação do potencial de transativação.

2.9. Quantificação da expressão dos novos genes da família NAC em condições de estresse abiótico e biótico

A avaliação da expressão dos genes da família NAC foi feita a partir do cDNA das plântulas de soja estressadas, conforme Tabela 3. Os genes avaliados e os oligonucleotídeos específicos utilizados encontram-se descritos na **Tabela 4**. Para cada reação, 20 ng de cDNA foram utilizados e a quantificação relativa da expressão de cada gene foi obtida pelo método de $2^{-\Delta\delta ct}$, avaliada em 3 amostras biológicas distintas (cada amostra é composta por um *pool* contendo 3 plantas fenotipicamente semelhantes), em duplicata, utilizando os protocolos recomendados pelo manual do kit SyBR Green® (Life Technologies).

Como gene normalizador, foi escolhido o gene da UKN-2 (Ma, *et al.*, 2013*) de soja e para confirmar o estado de estresse induzido por PEG, tunicamicina e ácido salicílico, utilizaram-se os genes marcadores *SMP*, *CNX* (calnexina), *PDI* (*protein disulfide isomerase*) e *PR-4* (NPR4). A amplificação seguiu o seguinte protocolo: 10 minutos a 95°C, seguido de 40 ciclos de amplificação a 94°C por 15 segundos e 60 °C por 1 minuto, seguido da etapa de dissociação.

(*) Shuhua Ma, Hongwei Niu, Chungi Liu, Jie Zhang, Chunyan Hou, Dongmei Wang (2013). Expression Stabilities of Candidate Reference Genes for RT-qPCR under Different Stress Conditions in Soybean. Plos One, e75271 – doi:10.1371/journal.pone.0075271

Tabela 4. Oligonucleotídeos iniciadores utilizados na quantificação da expressão dos genes NAC por qRT-PCR. Os oligonucleotídeos foram desenhados automaticamente pelo software PrimerExpress e foram escolhidos aqueles com penalidades mínimas.

Oligonucleotídeo	Sequência (5' - 3')	Código do Gene
qRT NAC85 fwd	TTCCAGGTAAGGCAGCGTTT	Glyma.12G149100
qRT NAC85 rev	TCGTGAACCATTCGGGTACTT	Glyma.12G149100
qRT NAC65 fwd	TGGGATTTGCCAGGTGATTT	Glyma.08G360200
qRT NAC65 rev	GAGCGATTTCCGTTGGGATA	Glyma.08G360200
qRT NAC072.A fwd	CTGGATGCCAACGAGAATC	Glyma.10G197600
qRT NAC072.A rev	TATCCCCATTCCCATTGCA	Glyma.10G197600
qRT NAC088.A fwd	GGGAGAGGACCGAATGGATTA	Glyma.12G186900
qRT NAC088.A rev	CTTGAGGCGACAAATGACCAA	Glyma.12G186900
qRT NAC119 fwd	AAATCCGGCAAGAGCAGAAG	Glyma.16G016400
qRT NAC119 rev	CCACTGCCCCGAAGATTTCA	Glyma.16G016400
UNK-2 fwd	GCCTCTGGATACCTGCTCAAG	
UNK-2 rev	ACCTCCTCCTCAAACCTCCTCTG	
SMP fwd	GCCGAACTGAGGAAAAGACGAACC	
SMP ver	CTTGGGCTGTTTGTGGGTCTTC	
PDI fwd	TTGGTTGAAGGCGTACAAGGATGG	
PDI ver	ACTCCAGCAGAACTATCTTCCCAG	
CNX fwd	TGATGGGGAGGAGAAGAAAAGGC	
PR-4 fwd	TGCGGGTGACAAATACAGGA	
PR-4 ver	TGCTGCACTGATCTACGATTCTC	

3. Resultados e Discussão

3.1. Análise *in silico* dos genes da superfamília NAC no genoma da soja

Os genes da família NAC vêm sendo estudados extensivamente por diversos grupos de pesquisa no mundo e, para a soja, a descrição mais completa da família foi fornecida por Le *et al.* (2011). Até então, 152 membros da superfamília (genes com ORFs completas) foram descritos no genoma da soja e 58 deles associados à resposta ao estresse hídrico [2].

Em nossa busca realizada na versão Willians82.v2.a2, disponibilizada pelo Phytozome em janeiro de 2016, 32 genes candidatos não descritos por Le *et al.* (2011) foram encontrados (**Tabela S1**). A busca contrastou o domínio NAC caracterizado no PFAM (código: PF02365) com as sequências de aminoácidos deduzidas de todo o genoma da soja e apenas genes que possuíam ORFs completas foram considerados como novos genes NAC. A Tabela S1 apresenta a nomenclatura Glyma.1 de todos os genes bem como a designação numérica, derivada da proposta por Le *et al.* (2011). A fim de manter a ordem numérica dos genes nos respectivos cromossomos foi introduzido uma letra para diferenciar os novos genes daqueles já descritos. Genes ainda não descritos foram destacados como *new* (destacados em amarelo na Tabela S1). Para cada gene, o tamanho da proteína correspondente foi mostrado, bem como as posições de início e fim do domínio NAC. Nessa análise, pudemos observar que 5 genes (destacados em laranja na Tabela S1) codificam para proteínas que possuem apenas o domínio NAC, que abrange até os 200 primeiros resíduos de aminoácidos. Sugere-se que esses genes sejam pseudo-genes ou que, se transcritos, possam expressar proteínas que atuem como repressores da família NAC, uma vez que esse domínio é responsável pela ligação ao DNA [3]. Como as proteínas não apresentam o domínio de ativação, supõe-se que elas sejam capazes de reconhecer *cis* elementos em promotores responsivos a fatores de transcrição do tipo NAC, sem contudo serem capazes de ativar a transcrição desses genes.

Dentre os genes já descritos, quatro não foram encontrados na nova anotação do genoma: GmNAC089, GmNAC096, GmNAC108 e GmNAC119 (destacados em verde claro na Tabela S1). Segundo informações disponibilizadas pelo Phytozome [®] (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax), para que os genes disponíveis na versão v1.1 pudessem receber um código de correspondência na versão Willians82.v2.a2, suas sequências de transcritos ou as sequências deduzidas de aminoácidos deveriam exibir acima de 90% de identidade. Com a melhoria das ferramentas de anotação de genomas, as sequências dos genes e seus *loci* são modificadas, fazendo com que predições de genes pertencentes a uma família possam ser modificadas a cada nova anotação do genoma. Dessa forma, a busca por genes da superfamília NAC em soja atualizou-a para 180 membros, sendo 148 já descritos e nomeados por Le *et al.* (2011) e 32 membros inéditos.

Quando a nomenclatura Glyma.1 foi proposta, ela otimizou a comunicação científica acerca da família NAC em soja devido a sua facilidade sistemática: os genes receberam como nome o número correspondente a ordem crescente de seus *loci* nos 20 cromossomos da soja, de 1 a 152 [2]. Buscando manter a nomenclatura Glyma.1, os novos genes foram organizados na **Tabela 5** e foi proposto um nome seguindo a ordem da distribuição dos NACs nos cromossomos e a ordem da nomenclatura Glyma.1. Nosso modelo propõe que o novo gene receba, como todos os demais, o prefixo *GmNAC*, o número do último gene descrito por Le *et al.* (2011) no *locus* mais próximo ao *locus* do novo gene e letras A, B e C subsequentemente. Exceção a esta regra refere-se aos novos genes Glyma.12G186900.1.p, Glyma.14G140100.1.p, e Glyma.16G016400.1.p que receberam os nomes GmNAC089, GmNAC108 e GmNAC119, respectivamente, substituindo os pseudo-genes anteriormente descritos por Le *et al.* (2011) em *loci* próximos.

Assim como predito por Le *et al.* (2011), que relatou a presença de 11 fatores de transcrição com domínios transmembrana, nossa análise também revelou a presença de 3 novos transfatores com domínio transmembrana (destacados em cinza na Tabela S1). A **Tabela 6** mostra os novos fatores de transcrição da família NAC que possuem domínio transmembrana, com a probabilidade do domínio NAC ser citosólico e a posição dos aminoácidos do domínio transmembrana. Precedentes na literatura demonstram a existência de fatores de transcrição NACs que possuem

domínio de ancoragem à membrana funcionais em outras espécies de plantas [35]. Esses transfatores são sintetizados em sua forma inativa e, apenas na presença de um sinal, sofrem proteólise limitada, sendo liberados de seu sítio de ancoragem e podendo desempenhar seu papel na reprogramação da expressão gênica no núcleo [35].

Tabela 5. Novos genes da superfamília NAC no genoma da soja. No contraste do domínio NAC contra a nova anotação do genoma da soja foram encontrados 32 genes NAC não mapeados por Le *et al.* (2011). Marcados em laranja encontram-se os possíveis pseudo-genes. Marcados em cinza encontram-se os transfatores que possuem domínio transmembrana.

Glyma v11.0	Le et al. 2011	Aminoácidos	Domínio NAC	E-value	Dom. Transm.
Glyma.02G050100.1.p	GmNAC006A	362	6 - 135	6.8E-53	
Glyma.02G284300.1.p	GmNAC013A	320	22 - 149	1.0E-47	
Glyma.03G164200.1.p	GmNAC014A	296	10 - 138	6.7E-48	
Glyma.03G179600.1.p	GmNAC014B	287	11 - 138	9.4E-48	
Glyma.04G014900.1.p	GmNAC015A	350	22 - 149	1.4E-49	
Glyma.04G175800.1.p	GmNAC017A	78	02 -- 78	1.3E-21	
Glyma.04G199000.1.p	GmNAC017B	169	28 - 73	4.0E-11	
Glyma.05G086000.1.p	GmNAC025A	177	11 - 140	1.5E-48	
Glyma.05G108700.1.p	GmNAC025B	212	18 - 98	6.4E-9	
Glyma.06G014900.1.p	GmNAC034A	374	22 -149	1.6E-49	
Glyma.06G288500.1.p	GmNAC044A	285	49 - 189	3.5E-30	
Glyma.07G047900.1.p	GmNAC046A	497	12 -141	2.1E-44	470 - 492
Glyma.07G192900.1.p	GmNAC050A	362	15 - 142	3.8E-51	
Glyma.08G163100.1.p	GmNAC059A	348	6 - 133	1.8E-54	
Glyma.10G077000.1.p	GmNAC071A	128	11 - 128	1.3E-38	
Glyma.10G077400.1.p	GmNAC071B	136	11 - 136	1.7E-43	
Glyma.10G197600.1.p	GmNAC072A	448	5 -- 97	2.1E-29	424 - 446
Glyma.10G204700.1.p	GmNAC072B	422	56 - 195	1.5E-28	
Glyma.12G003200.1.p	GmNAC080A	356	9 - 138	3.2E-51	
Glyma.12G145100.1.p	GmNAC083A	180	11 - 138	1.6E-46	
Glyma.12G160100.1.p	GmNAC085A	133	2 - 115	5.2E-35	
Glyma.12G186900.1.p	GmNAC089	493	17 - 143	8.8E-47	303 - 325
Glyma.13G294000.1.p	GmNAC102A	279	48 - 188	6.1E-30	
Glyma.14G140100.1.p	GmNAC108	373	27 - 154	4.8E-53	
Glyma.16G016400.1.p	GmNAC119	267	11 - 140	1.0E-44	
Glyma.16G069300.1.p	GmNAC125A	399	56 - 198	4.8E-27	
Glyma.18G301500.1.p	GmNAC139A	229	14 - 138	1.3E-49	
Glyma.19G002900.1.p	GmNAC139B	389	56 - 197	1.9E-28	
Glyma.19G108800.1.p	GmNAC143A	336	8 - 138	5.8E-51	
Glyma.19G165600.1.p	GmNAC144A	294	10 - 138	4.1E-48	
Glyma.19G195800.1.p	GmNAC145A	254	4 - 134	7.5E-43	
Glyma.20G175500.1.p	GmNAC149A	341	19 - 147	1.6E-51	

Tabela 6. Novos fatores de transcrição da família NAC com domínio transmembrana.

Glyma v11.0	Le et al. 2011	Aminoácidos	Domínio NAC	E-value	Dom. Transm.	Prob.*
Glyma.07G047900.1.p	GmNAC046A	497	12 -141	2.1E-44	470 - 492	21,93%
Glyma.10G197600.1.p	GmNAC072A	448	5 -- 97	2.1E-29	424 - 446	0,40%
Glyma.12G186900.1.p	GmNAC089	493	17 - 143	8.8E-47	303 - 325	9,86%

(*) Probabilidade relativa à posição do domínio NAC – análises feitas pelo software TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) indicam a probabilidade da porção N-terminal, nesse caso, correspondente ao domínio NAC, estar voltada para o citosol.

Apesar do mecanismo de ação dos transfatores solúveis e dos transfatores associados à membrana serem os mesmos, a localização celular e a estrutura das proteínas nos permite dividir a superfamília NAC em 2 grandes grupos: (i) grupo de transfatores do tipo NAC solúveis, composto por 166 membros, cuja localização celular supõe-se ser nuclear; (ii) grupo de transfatores do tipo NAC com domínio transmembrana, composto por 14 membros, cuja localização celular é associada às membranas biológicas, tais como a membrana plasmática ou a membrana do retículo endoplasmático.

O domínio NAC é bastante conservado em sequência e estrutura, sendo dividido em 5 regiões de A-E características em todos os membros da família [2-4]. A **Figura 2** mostra o alinhamento múltiplo das sequências dos 200 primeiros aminoácidos, compreendendo o domínio NAC das novas proteínas com a sequência de GmNAC30, já caracterizado por nosso grupo de pesquisa [5]. Pode-se observar que as cinco regiões (A-E) são altamente conservadas. Os resíduos marcados em vermelho representam a maior frequência desse resíduo na referida posição dentre todos os novos membros e os resíduos marcados em azul representam a segunda maior frequência. Dentro de cada subdomínio, existem resíduos com 100% de conservação (destacados em negrito) ou muito próximo disso, como os resíduos de glicina nas regiões A e D, prolina nas regiões A e B, resíduos de triptofano nas regiões C e D, os resíduos de glutamato na região A e os resíduos de arginina na região C.

A **Figura 3** mostra o remapeamento de todos os genes NAC nos cromossomos da soja.

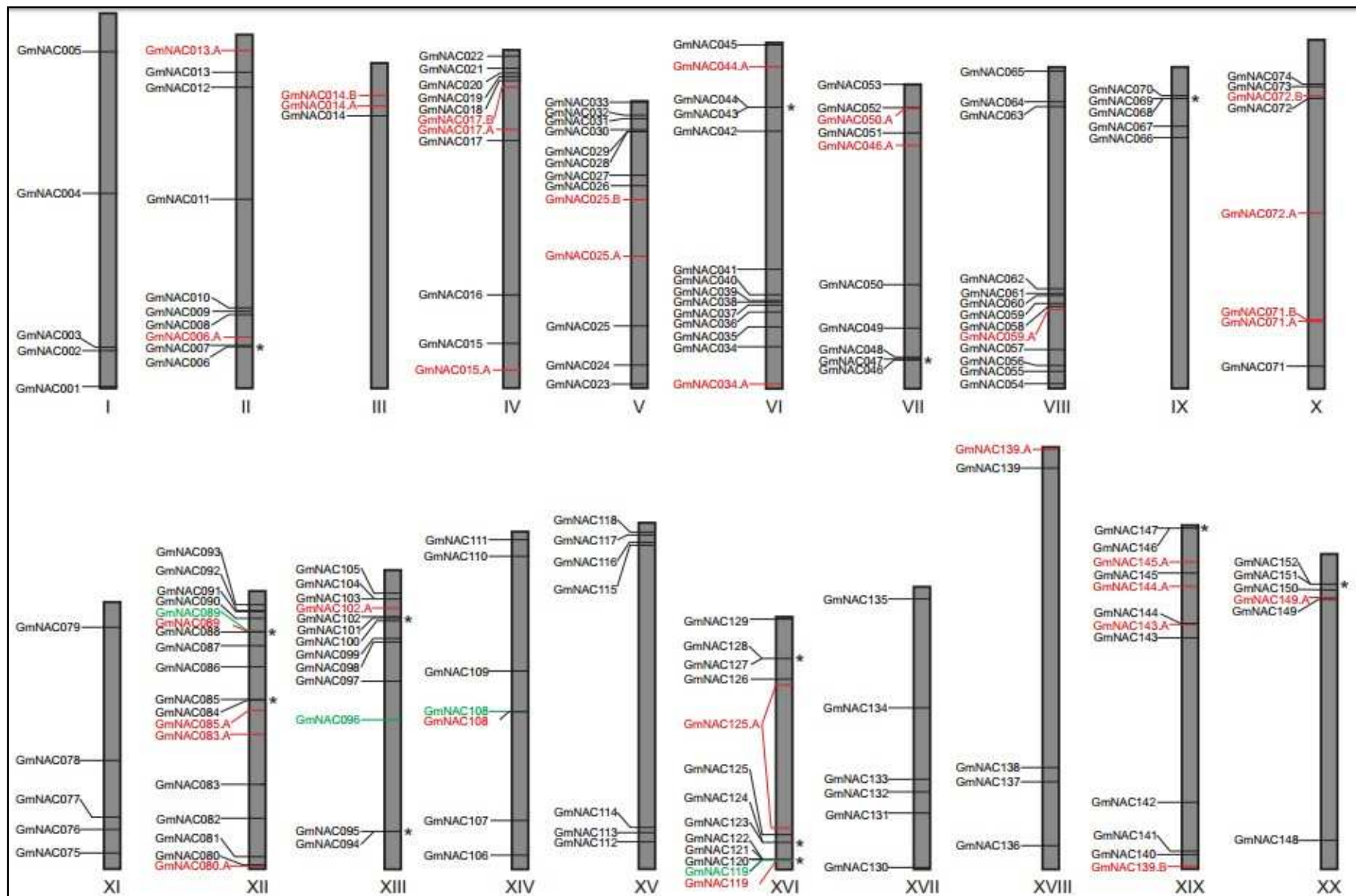


Figura 3. Distribuição dos genes da superfamília NAC nos cromossomos da soja. Em preto, encontram-se os genes anteriormente descritos por Le et al. (2011) e mapeados de acordo com seus loci na nomenclatura Glyma.1. Destacados em vermelho, os novos genes. Em verde, encontram-se os genes que foram descritos por Le et al. (2011), mas não foram encontrados na revisitação do genoma. Os (*) referem-se a blocos de repetição em tandem.

Os genes já mapeados por Le *et al.* (2011) encontram-se nas mesmas posições definidas anteriormente. Os genes GmNAC089, GmNAC096, GmNAC108 e GmNAC119, que não foram encontrados na nova anotação do genoma, estão destacados em verde. Em vermelho, encontram-se os novos genes. Os asteriscos à direita dos cromossomos indicam possíveis genes duplicados em blocos em tandem. O cromossomo 12 foi mapeado com maior número de genes NAC, contendo 14 dos 152 membros anteriormente descritos [2]. Após o remapeamento, essa observação ainda é verdadeira, sendo que 4 novos genes NAC foram mapeados no cromossomo 12, totalizando 17 membros dos 180 totais. Os cromossomos 10 e 19 também receberam 4 genes novos no remapeamento, enquanto a maioria dos demais cromossomos recebeu apenas 1 ou 2 novos genes. A **Figura 4** apresenta um gráfico de setores com a porcentagem de genes da família NAC em cada cromossomo da soja após o remapeamento.

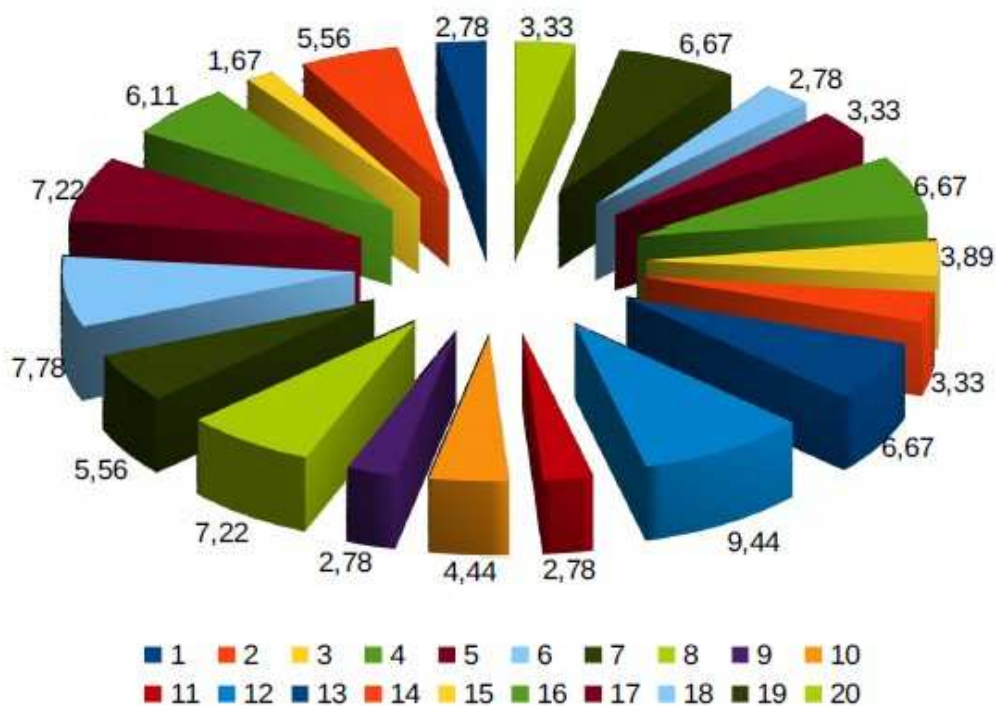


Figura 4. Distribuição percentual dos genes da família NAC nos cromossomos da soja. O gráfico de setores indica qual a porcentagem de genes NAC (valor superior) distribuídos nos 20 cromossomos da soja.

A fim de examinar a estrutura da família NAC de soja, procedeu-se à uma análise filogenética das sequências deduzidas de aminoácidos dos genes de soja junto com os genes NACs de *Arabidopsis* (**Figura 5**). Conforme, ilustrado na Figura 4, os genes foram classificados em sub-famílias e sub-grupos juntos aos seus ortólogos de *Arabidopsis*, sendo alguns deles já funcionalmente caracterizados, o que norteou a inferência de algumas funções possivelmente associadas aos genes de soja. Nossa análise classificou os genes NAC de soja em 5 grandes subfamílias, sendo a subfamília NAP (*NAC-like activated by APETALA/PISTILLATA*), VND (*vascular-related NAC domain*), TIP (*Turnip crinkle virus interaction protein like*), CUC (*cup-shaped cotyledon*) e ATAF (*Arabidopsis transcription fator like*), destacadas em amarelo, laranja, azul escuro, verde e vermelho, respectivamente. Os novos genes encontrados em nossa revisitação ao genoma da soja encontram-se destacados em vermelho, com destaque em rosa para os possíveis pseudo-genes encontrados. Em azul, encontram-se os genes que foram clonados no presente trabalho.

Os membros da subfamília CUC (destacada em verde) tem uma relevância histórica no estudo dos fatores de transcrição do tipo NAC. Esse grupo foi um dos primeiros a serem estudados e correlacionados à manutenção dos meristemas e separação dos cotilédones em *Arabidopsis thaliana*. Plantas mutantes para esse gene apresentam um fenótipo aberrante do órgão [8-9] e não desempenham suas funções fisiológicas normalmente. Os fatores de transcrição da família NAC, apesar de serem extensivamente estudados como alvos no entendimento das respostas aos estresses ambientais, também podem desempenhar papéis importantes na regulação da morfofisiologia do vegetal. Portanto, baseado no agrupamento filogenético e sua consistência/estabilidade (revelada pelos valores de bootstrap de cada clado), podemos dizer que o genoma da soja conta com uma vasta variedade de fatores de transcrição do tipo NAC que possuem relação com a manutenção da fisiologia normal do vegetal. Dentro dessa subfamília, 7 novos genes foram encontrados (destacados em vermelho), sendo os pares *Glyma.03G164200* e *Glyma.19G165600* e *Glyma.06G014900* e *Glyma.04G014900* intimamente relacionados entre si.

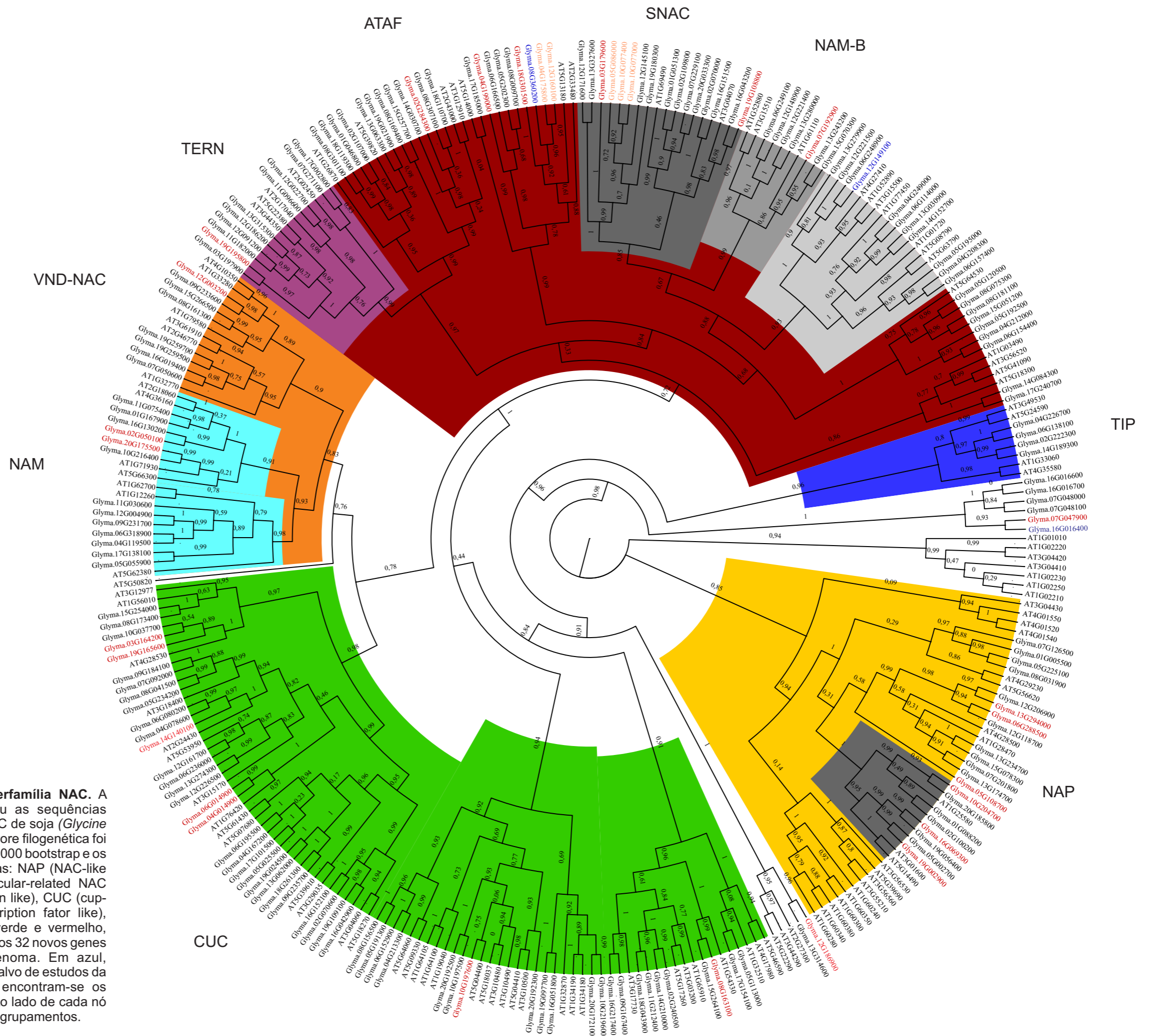


Figura 5. Agrupamento filogenético da superfamília NAC. A construção do agrupamento filogenético reuniu as sequências deduzidas de aminoácidos de todos os genes NAC de soja (*Glycine max*) e *Arabidopsis thaliana*. A reconstrução da árvore filogenética foi feita pelo software Fasttree em um consenso de 10.000 bootstraps e os genes foram agrupados em 5 grandes subfamílias: NAP (NAC-like activated by APETALA/PISTILLATA), VND (vascular-related NAC domain), TIP (Turnip crinkle virus interaction protein like), CUC (cup-shaped cotyledon) e ATAF (Arabidopsis transcription fator like), destacadas em amarelo, laranja, azul escuro, verde e vermelho, respectivamente. Destacados em vermelho, estão os 32 novos genes candidatos encontrados na reavaliação do genoma. Em azul, encontram-se os genes que foram clonados como alvo de estudos da resposta aos estresses ambientais. Em rosa, encontram-se os possíveis pseudo-genes revelados pela análise. Ao lado de cada nó encontram-se os valores de bootstrap obtidos nos agrupamentos.

Outra grande subfamília com destaque histórico é a subfamília ATAF (destacada em vermelho). Dentro da subfamília, os genes são intimamente relacionados e já se tem demonstrado sua relação com as respostas aos estresses bióticos: a superexpressão de *ATAF-2* aumenta a suscetibilidade de plantas a fungos necrotróficos, como *Fusarium oxysporum*, e reduz a incidência de sintomas de infecção de tabaco por TMV (*Tobacco mosaic virus*) devido ao controle da expressão de genes PR [28-31]. *ATAF-1* promove resistência a penetração de fungos biotróficos, como *Blumeria graminis* e, paralelamente, aumenta a suscetibilidade a *Pseudomonas syringae*, *Botrytis cinerea* e *Alternaria brassicicola* [30, 36].

Dentro da subfamília ATAF, subgrupos podem ser formados de acordo com a função biológica de alguns membros caracterizados em soja ou ortólogos em *Arabidopsis*. Em cinza claro, encontra-se o grupo de genes da subfamília ATAF relacionados a *ANAC019*. Esses genes constituem um grupo pequeno, cuja função permeia a resposta a estresses abióticos, sendo induzidos por ABA, seca e alta salinidade, além do fato de que plantas que superexpressam genes desse grupo apresentam-se mais tolerantes ao estresse hídrico [37]. Em cinza médio, encontra-se destacado o grupo dos genes NAM-B (*no apical meristhem*), relacionados à regulação do processo de senescência e resposta ao estresse hídrico, além da regulação do desenvolvimento de meristemas. Ainda dentro da subfamília ATAF, tem-se o grupo SNAC (destacado em cinza escuro - também presente dentro da subfamília NAP), relacionado as respostas aos estresses, principalmente aquelas coordenadas pelo ácido jasmônico. De maneira geral, os genes classificados como SNAC regulam as vias de resistência a patógenos, resposta a injúrias mecânicas e ataque de pragas, sendo coordenados por diversos sinais hormonais como jasmonato, ácido salicílico e etileno [11-16, 28-30].

Os genes do grupo TERN (*tobacco elicitor-responsive encoding NAC-domain protein*), destacados em roxo e da subfamília TIP, destacada em azul escuro, executam um papel biológico comum, relacionado à resposta induzida por patógenos. Conforme vimos, TIP é uma proteína que interage com o capsídeo viral do *Turnip crinkle virus* e desencadeia uma hipersensibilização na planta e o gene *TERN* é induzido por elicitores da resposta a patógenos [28-32]. Dentro desses grupos, alguns genes controlam a morte celular programada em situações de estresse de seca e de retículo endoplasmático. Dentre esses genes, encontra-se *GmNAC81* (Glyma.12G022700.1.p), que possui apenas um gene ortólogo em

Arabidopsis, ANAC36 (AT2G17040.1) e, consistente com uma função em morte celular, agrupa-se com os membros do grupo TERN. A proteína codificada por esse gene, em um complexo com GmNAC30 (que pertence à família ATAF), induz a expressão da proteína VPE, que executa a morte celular programada [5, 11, 32-33].

Na nova varredura, 12 novos genes foram colocados na subfamília ATAF, sendo que *GmNAC65* e *GmNAC85* foram clonados como alvos para o estudo das respostas aos estresses ambientais e tiveram sua expressão avaliada em diferentes condições de estresse em plântulas de soja por PCR quantitativo em tempo real.

As subfamílias NAP (destacada em amarelo) e VND (destacada em laranja) estão relacionadas ao controle da senescência foliar [36-37] e com a formação de vasos do xilema em *Arabidopsis thaliana* [38]. Dentro da subfamília VND, existe um grupo de proteínas relacionadas ao grupo NAM. Dentro desses grupos, 9 novos genes foram encontrados.

3.2. Clonagem de *GmNAC065*, *GmNAC085* e *GmNAC119*

Como alvos de respostas adaptativas aos estresses ambientais, as ORFs de 3 genes da família NAC foram clonados em vetores de entrada do sistema Gateway® (Invitrogen): *GmNAC065* (Glyma.08G360200), *GmNAC085* (Glyma.12G149100) e *GmNAC119* (Glyma.16G016400). Os fatores de transcrição *GmNAC065* e *GmNAC085* foram previamente caracterizados por Le *et al.* (2011) e pertencem à subfamília ATAF, cuja principal característica é seu forte envolvimento nas respostas aos estresses abióticos, como a seca e a salinidade. Recentemente, foi demonstrado por meio de hibridizações por microarranjo de DNA que o gene *GmNAC085* é fortemente induzido por déficit hídrico, enquanto que o gene *GmNAC065* é negativamente regulado por seca, o que fundamentou a escolha desses dois genes para clonagem e posterior análise funcional [14]. A amplificação do cDNA dos respectivos genes a partir de bibliotecas de cDNAs preparadas de folhas em diferentes condições de estresses, revelou que, de fato, *GmNAC085* é expresso e induzido em condições de estresse hídrico, sendo detectado em cDNAs de plântulas de soja estressadas com PEG por até 12 horas, enquanto que *GmNAC065* não foi detectado nas mesmas condições, mas mostrou-se expresso em plântulas submetidas ao estresse de retículo endoplasmático elicitado por tunicamicina (dados não mostrados). Essa possível divergência no papel regulatório entre *GmNAC065* e *GmNAC085*, que pertencem à subfamília ATAF, é interessante do ponto de vista da caracterização funcional de transfatores da superfamília NAC, que exibem grande plasticidade de função apesar da estrutura do domínio ser altamente conservada. As **Figuras 6 e 7** representam a racionalização da clonagem das ORFs dos genes *GmNAC065* e *GmNAC085*, com as respectivas sequências de cDNA e aminoácidos, destacando as regiões conservadas de A-E do domínio NAC. O diagnóstico em gel de agarose para a purificação do fragmento amplificado está representado nas **Figuras S1 e S2**.

A revisitação do genoma da soja revelou a presença de 32 novos genes que não haviam sido descritos no inventário da superfamília NAC feito por Le *et al.* (2011). Como alvo dentre esses novos genes, *GmNAC119* (Glyma.16G016400) também foi clonado pelo sistema Gateway®. A proteína NAC119 pertence a um ramo filogenético distinto das subfamílias mais bem caracterizadas, como a subfamília ATAF e CUC, mas está em um ramo entre as subfamílias NAP e TIP, que

1	atggagaaggtgaat tttgtgaagaatggagagcttagattgcct	
	M E K V N F V K N G E L R L P	(A)
46	ccggggttccg tttccacccgactgatgaggagctggttctgcaa	
	P G F R F H P T D E E L V L Q	
91	tacttgaagcgc aaggtcttctcctgccctttgccagcctctatc	
	Y L K R K V F S C P L P A S I	(B)
136	attcctgagctt catgtttgcaagtctgatccttgggatttgcca	
	I P E L H V C K S D P W D L P	
181	ggtgatttggag caagagagatacttcttttagcaccaaagtggcc	
	G D L E Q E R Y F F S T K V A	(C)
226	aaatatcccaac ggaaatcgctccaacagagccacaaattcgggt	
	K Y P N G N R S N R A T N S G	
271	tattggaaggca actggcttggacaaacaaattggttacttcaaaa	
	Y W K A T G L D K Q I V T S K	
316	ggcaacaacca agttgtcggaatgaagaagacacttgtttctac	
	G N N Q V V G M K K T L V F Y	
361	agaggcaagcct cctaatggatccagaactgattggatcatgcac	
	R G K P P N G S R T D W I M H	(D)
406	gagtatcgctcat ccttaacgcctctcagtctcagagccatggt	
	E Y R L I L N A S Q S Q S H V	
451	gttcccatggaaa attgggttctgtgtcgcata tttttgaagagg	
	V P M E N W V L C R I F L K R	(E)
496	agaattggtgct aaaaaatggggaggagagcaactctaaggtggtt	
	R I G A K N G E E S N S K V V	
541	ttctatgacttct agcgcagaacaagaccgattcctcctcatcg	
	F Y D F L A Q N K T D S S S S	
586	gtcgcagtggaat tacacatgaatcagatgaacatgaagagagc	
	V A S G I T H E S D E H E E S	
631	agtagctccaacac cttcccttatactattagaagaaaaccttaa 675	
	S S S N T F P Y T I R R K P *	

Figura 6. Sequência de aminoácidos e nucleotídeos da ORF de *GmNAC065*. O gene *GmNAC065* (Glyma.08G360200) foi clonado à partir de uma mistura de cDNAs de plântulas de soja sob diferentes condições de estresse. Nas linhas enumeradas, encontra-se a sequência de cDNA do gene disponível no Phytozome ® e abaixo de cada linha enumerada os aminoácidos codificados pelos códons da ORF. Em destaque, encontram-se as regiões conservadas do domínio NAC.

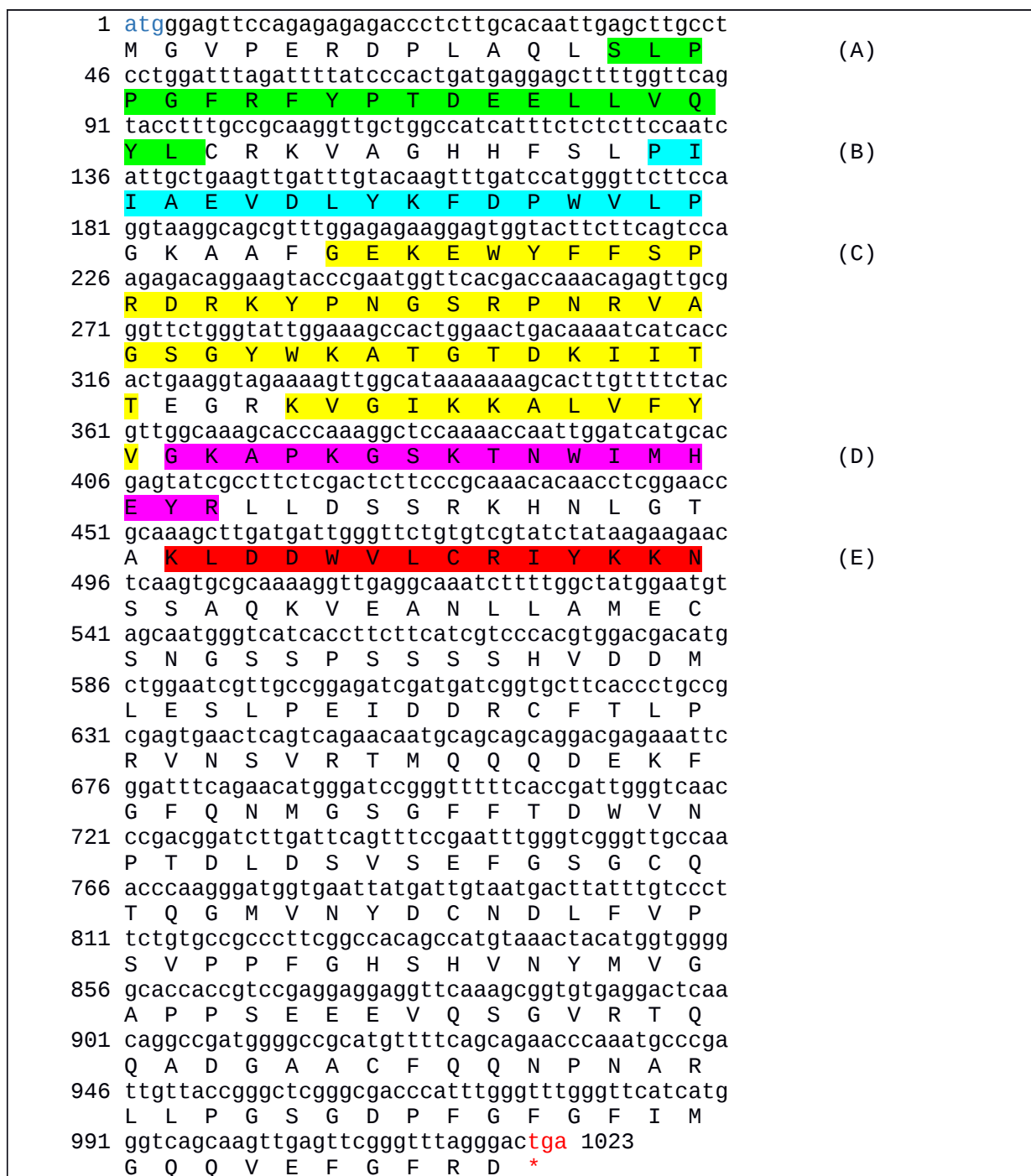


Figura 7. Sequência de aminoácidos e nucleotídeos de *GmNAC085*. O gene *GmNAC085* (Glyma.12G149100) foi clonado à partir de uma mistura de cDNAs de plântulas de soja sob diferentes condições de estresse. Nas linhas enumeradas, encontra-se a sequência de cDNA do gene disponível no Phytozome ® e abaixo de cada linha enumerada os aminoácidos codificados pelos códons da ORF. Em destaque, encontram-se as regiões conservadas do domínio NAC.

desempenham funções relacionadas ao controle da senescência foliar e à resposta aos estresses bióticos [36-37]. A **Figura 8** mostra a sequência de aminoácidos e nucleotídeos da ORF de *GmNAC119* e o diagnóstico em gel de agarose da amplificação e purificação do fragmento amplificado encontra-se representado na **Figura S3**.

1	atggagaacataatttccatatgtgatcacatgcctgtaggattc	
	M E N I I S I C D H M P V G F	(A)
46	agattccgtcccacagatgaagagcttgtcgactattacctcaa	
	R F R P T D E E L V D Y Y L K	
91	cacaagcttctggctgatgatttccctgtccacatcatccctgaa	
	H K L L A D D F P V H I I P E	(B)
136	atcgacctttgcaaggtggaaccttgggacgtaccaggaagatca	
	I D L C K V E P W D V P G R S	
181	gtgatcaaatccgatgaccagaatggtttttcttcagtcctgtg	
	V I K S D D P E W F F F S P V	(C)
226	gattacaagtacttgaagagtaaaggtttaacaggacaaccaag	
	D Y K Y L K S K R F N R T T K	
271	cgtggcttctggaaagccactgggaatgaccgaaaataaggata	
	R G F W K A T G N D R K I R I	
316	ccgggcaccagcaatgtcattgggactaagaagactcttgtttc	
	P G T S N V I G T K K T L V F	(D)
361	caccaaggccgtgttccccgtggtgccaagaccaactgggttatt	
	H Q G R V P R G A K T N W V I	
406	catgaataccatgctgttacctcccataagccagagggtttt	
	H E Y H A V T S H E S Q R A F	(E)
451	gttttgtgccgcttgatgaagaaagctgagaaaaaaatgaagg	
	V L C R L M K K A E K K N E G	
496	ggaattgaggcaccgtcttttgatgaaggggagcccagtatacac	
	G I E A P S F D E G E P S I H	
541	atggtttctgactatggaaatccggcaagagcagaaggcattgca	
	M V S D Y G N P A R A E G I A	
586	tctgtagatattattcctggagtagatatggatgaaatctttcgg	
	S V D I I P G V D M D E I F R	
631	gcagtggatcaagctgagaaattctccccccagtacaacagtct	
	A V D Q A E K F S P P V Q Q S	
676	ccaactggcattgggctggaagatTTTTTCCAAACTCTCATTG	
	P T G I G L E D F F P N S P L	
721	ttcaatgcccgttttggaaagtgaaaacatcaatatgcaaacttca	
	F N A R F G S E N I N M Q T S	
766	cttgaagttacggatgacgaagatgagttcataccctag 804	
	L E V T D D E D E F I P *	

Figura 8. Sequência de aminoácidos e nucleotídeos de *GmNAC119*. O gene *GmNAC119* (Glyma.16G016400) foi clonado à partir de uma mistura de cDNAs de plântulas de soja sob diferentes condições de estresse. Nas linhas enumeradas, encontra-se a sequência de cDNA do gene disponível no Phytozome ® e abaixo de cada linha enumerada os aminoácidos codificados pelos códons da ORF. Em destaque, encontram-se as regiões conservadas do domínio NAC.

Todos os clones foram enviados ao sequenciamento automatizado de Sanger (Myleus Biotecnologia - <http://www.myleus.com/>) para a confirmação de suas identidades. Os resultados dos alinhamentos gerados confirmaram a identidade de cada ORF clonada e encontram-se nas **Figuras S4-S6**.

3.3. Localização Celular de NAC119

Os fatores de transcrição são proteínas celulares que desempenham sua função no núcleo. Apesar de alguns deles apresentarem domínios transmembrana, a maioria dos transfatores são solúveis e, depois de expressos no citosol, são importados pela maquinaria de importação nuclear, onde irão reconhecer os *cis* elementos regulatórios de genes alvo. A **Figura 9** mostra a predição do sinal de localização nuclear de NAC119 (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi), que revela a presença de 3 regiões que são reconhecidas pela maquinaria de importação nuclear.

Predicted NLSs in query sequence	
MENIISICDHMPVGFRRPTDEELVDYLLKHKLLADDFPVHIPEIDLCK	50
VEPWDPGRSVIKSDDPEWFFSPVDYKYLKSKRFNRTTKRGFWKATGND	100
RKIRIPGTSNVIGTKKTLVFHQGRVPRGAKTNWVIHEYHAVTSHESQRAF	150
VLRLMKAEEKKNEGGIEAPSFDEGEPSIHMVSDYGNPARAEGIASVDII	200
PGVDMDEIFRAVDQAEKFSPPVQQSPTGIGLEDFFPNSPLFNARFGSENI	250
NMQTSLEVTDEDEFIP	267

Predicted bipartite NLS		
Pos.	Sequence	Score
55	DVPGRSVIKSDDPEWFFSPVDYKYLKSKRF	3.1
99	NDRKIRIPGTSNVIGTKKTLVFHQGRVPRG	3.2
154	RLMKAEEKKNEGGIEAPSFDEGEPSIHMVSDY	3.2

Figura 9. Predição do sinal de localização nuclear de NAC119. A proteína NAC119 possui um sinal de localização nuclear de proteínas que não são exclusivamente nucleares (score 3.0). No quadro superior, encontra-se a sequência de aminoácidos de NAC119, com o sinal tripartido destacado em vermelho. No quadro inferior, encontram-se as posições de cada segmento do sinal de localização nuclear e os scores para cada região.

Essa predição apresenta um *score* de 3.0, que compreende proteínas que não são exclusivamente nucleares (scores superiores a 5.0, de acordo com os parâmetros que modelam o software – disponível em http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS Mapper_help.cgi#score). No entanto, como esses fatores de transcrição fazem parte de redes de sinalização celular e somente são enviados ao núcleo frente a um sinal específico na resposta da planta em diferentes situações, pode-se considerar que o sinal encontrado realmente é capaz de promover a importação de NAC119 para o núcleo.

Para confirmar a predição *in silico* da localização de NAC119, a construção NAC119-GFP (clone pUFV 2830) em agrobactéria (linhagem GV3101) foi inoculada em folhas de *Nicotiana benthamiana* e as plantas foram levadas ao microscópio confocal de varredura. A **Figura 10 A** mostra intensa fluorescência na região nuclear, indicado pela seta na micrografia de campo claro – **Figura 10 B**. Estes resultados demonstram que a proteína GmNAC119 está localizada no núcleo da célula vegetal em condições normais.

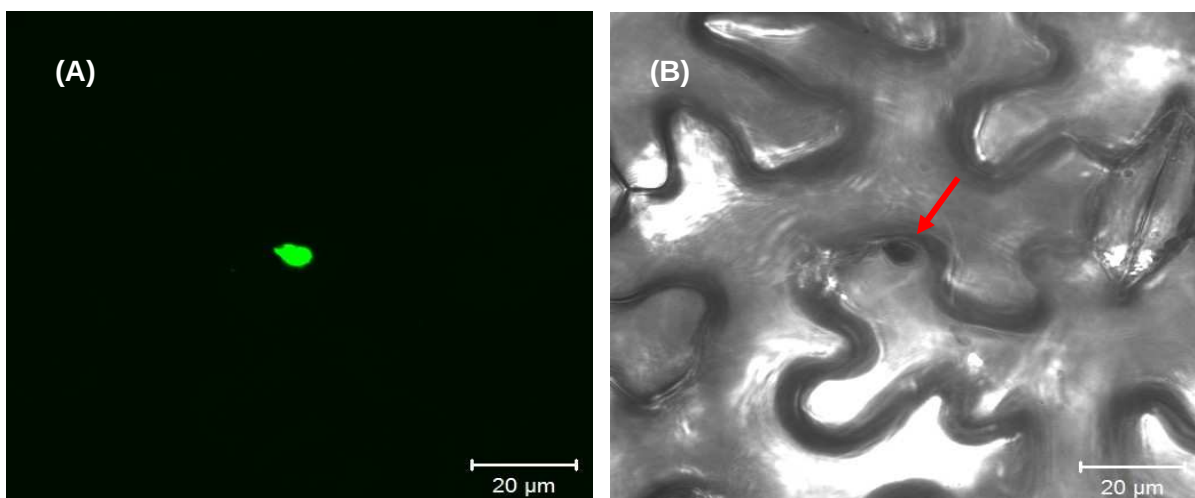


Figura 10. Localização celular de NAC119. A proteína NAC119 foi fusionada à porção aminoterminal da proteína fluorescente GFP e agroinfiltrada em folhas de *Nicotiana benthamiana*. Em (A) pode-se observar a fluorescência da construção NAC119-GFP em uma região restrita ao núcleo, sem nenhum espalhamento para outras regiões celulares. Em (B), a seta vermelha aponta o núcleo na micrografia de campo claro da região fluorescente correspondente.

3.4. Avaliação da capacidade de transativação de NAC119 em sistema de duplo híbrido em leveduras

Fatores de transcrição, na maioria das vezes, interagem entre si na forma de dímeros para que possam realizar o reconhecimento de elementos regulatórios na região promotora de genes responsivos àqueles transfatores e ativar a expressão gênica. Alguns deles, como já demonstrado para NACs em soja [32], podem formar homodímeros e são capazes de transativarem a expressão de genes sem a necessidade de outros fatores de transcrição ou proteínas acessórias.

Para avaliar a atividade transativadora de GmNAC119, a proteína fusionada ao domínio de ligação de GAL4 (clone pUFV 2828) foi transformada em leveduras (linhagem AH 109). A **Figura 11** demonstra que NAC119 não é capaz de transativar, pois não exibiu qualquer crescimento significativo, tal como o vetor vazio. Apenas na ausência de 3-AT em meio SD/-leu/-his/ foi possível detectar uma única colônia nas gotas de densidade ótica 1,0 e 0,1, mostrando que o fator de transcrição candidato não reporta a atividade de *HIS3*. Em nenhuma das concentrações crescentes de 3-AT foi possível identificar qualquer crescimento de transformantes NAC119, revelando um padrão idêntico ao controle negativo (pDEST32 vetor vazio). Como controle positivo do protocolo de transformação da levedura, apresenta-se o crescimento de transformantes com pDEST32 vazio e NAC119 em meio SD deficiente apenas em leucina.

3.5. Análise do perfil de expressão gênica sob condições de estresse

Os fatores de transcrição da superfamília NAC controlam processos de desenvolvimento, diferenciação de órgãos e tecidos vegetais e também são responsivos a situações de adversidades ambientais, como os estresses biótico e abiótico. A expressão de genes que codificam fatores de transcrição exibe os mais diferentes perfis no que tange o momento fisiológico do vegetal, variando de acordo com o estágio de desenvolvimento das plantas, o tecido alvo, as condições de crescimento/desenvolvimento e outros fatores.

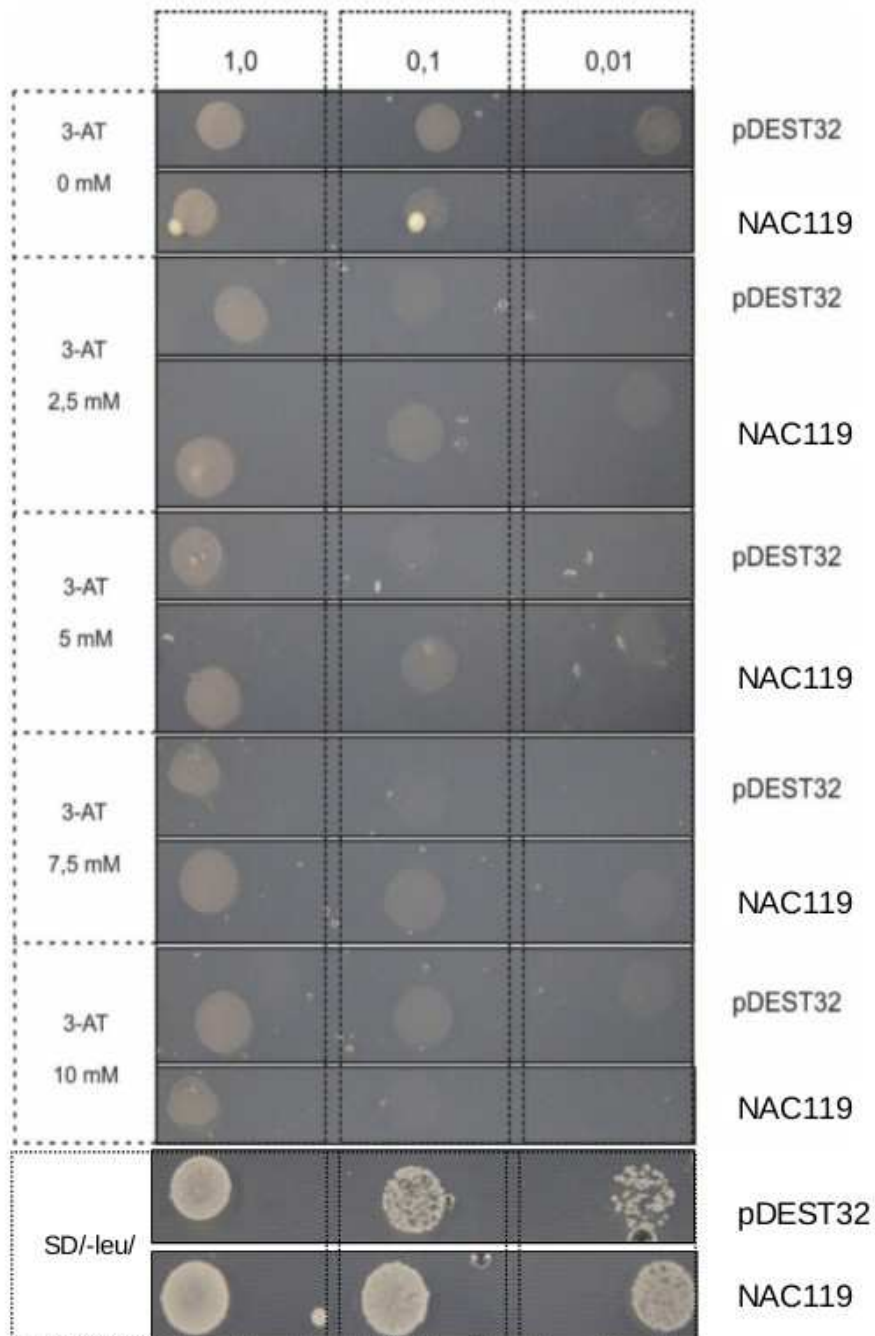


Figura 11. Ensaio de transativação de NAC119 em sistema de duplo híbrido em leveduras. A construção NAC119-BD(GAL4) e o vetor vazio pDEST32 foram transformados em leveduras pelo protocolo PEG-acetato de lítio. Depois de selecionados, os transformantes foram crescidos até a densidade ótica superior a 1,0 e plaqueados em meio SD/-leu/-his/. Foram feitas diluições da cultura com densidades óticas de 1,0, 0,1 e 0,01, e o potencial de transativação testado na presença de 3-aminotriazol em concentrações crescentes de 0 mM – 10 mM. As placas foram incubadas durante 6 dias a 28 °C. O quadro abaixo refere-se ao controle (+) da transformação, com as leveduras plaqueadas em meio SD/-leu/.

Para validar a existência dos novos genes NAC preditos pela análise *in silico* e correlacioná-los com seus grupos filogenéticos foram escolhidos os genes *GmNAC065* (já descrito por Le *et al.* (2011)), *GmNAC072.A*, *GmNAC089* e *GmNAC119*, que são novos genes preditos na revisitação do genoma da soja, para serem avaliados quanto aos seus perfis de expressão gênica por qRT-PCR em resposta a diferentes estresses. As plântulas de soja foram estressadas nos tempos de 0,5 hora, 2 horas e 12 horas com PEG, tunicamicina e ácido salicílico, que são tratamentos que simulam estresse osmótico, estresse no retículo endoplasmático e estresse biótico, respectivamente. Para monitorar a eficiência dos tratamentos, foram utilizados como marcadores os genes *SMP* (*seed maturation protein*), *CNX* (calnexina), *PDI* (*protein disulfide isomerase*) e *PR-4* (*NPR4*). O gene *SMP* é marcador do estresse osmótico, enquanto os genes *PDI* e *CNX* são marcadores do estresse no retículo endoplasmático e o gene *PR-4* é induzido em condições de estresse biótico.

As plantas respondem aos estímulos ambientais por meio de cinéticas de resposta, as quais podem ser classificadas como precoce, intermediária e tardia. Visando estabelecer a cinética de resposta dos genes alvos selecionados, escolheram-se os tempos de 0,5 h, 2 h e 12 h, abrangendo a faixa de cinética precoce, intermediária e tardia, respectivamente. Os gráficos no topo das **Figuras 12-14** comprovam o sucesso dos tratamentos de estresse nas plantas por meio da expressão diferencial dos genes marcadores em pelo menos um dos pontos da cinética.

A **Figura 12** mostra o perfil da expressão dos genes alvo em folhas de plântulas de soja estressadas com PEG. Dentre os genes avaliados, nenhum apresentou uma expressão diferencial estatisticamente significativa. O gene *GmNAC119* apresentou uma tendência de aumento de sua expressão a partir do tempo 2 horas, no entanto, com esse desenho experimental não foi possível encontrar uma diferença estatisticamente significativa em relação ao controle (CTR). No caso do gene *GmNAC65*, o qual já foi mostrado ser negativamente regulado em condições de estresse hídrico severo [14], também não mostrou nenhuma diferença significativa de expressão nos tempos testados. Esses dados sustentam a hipótese de que, apesar da alta conservação de sequência e de estrutura do domínio NAC, seus

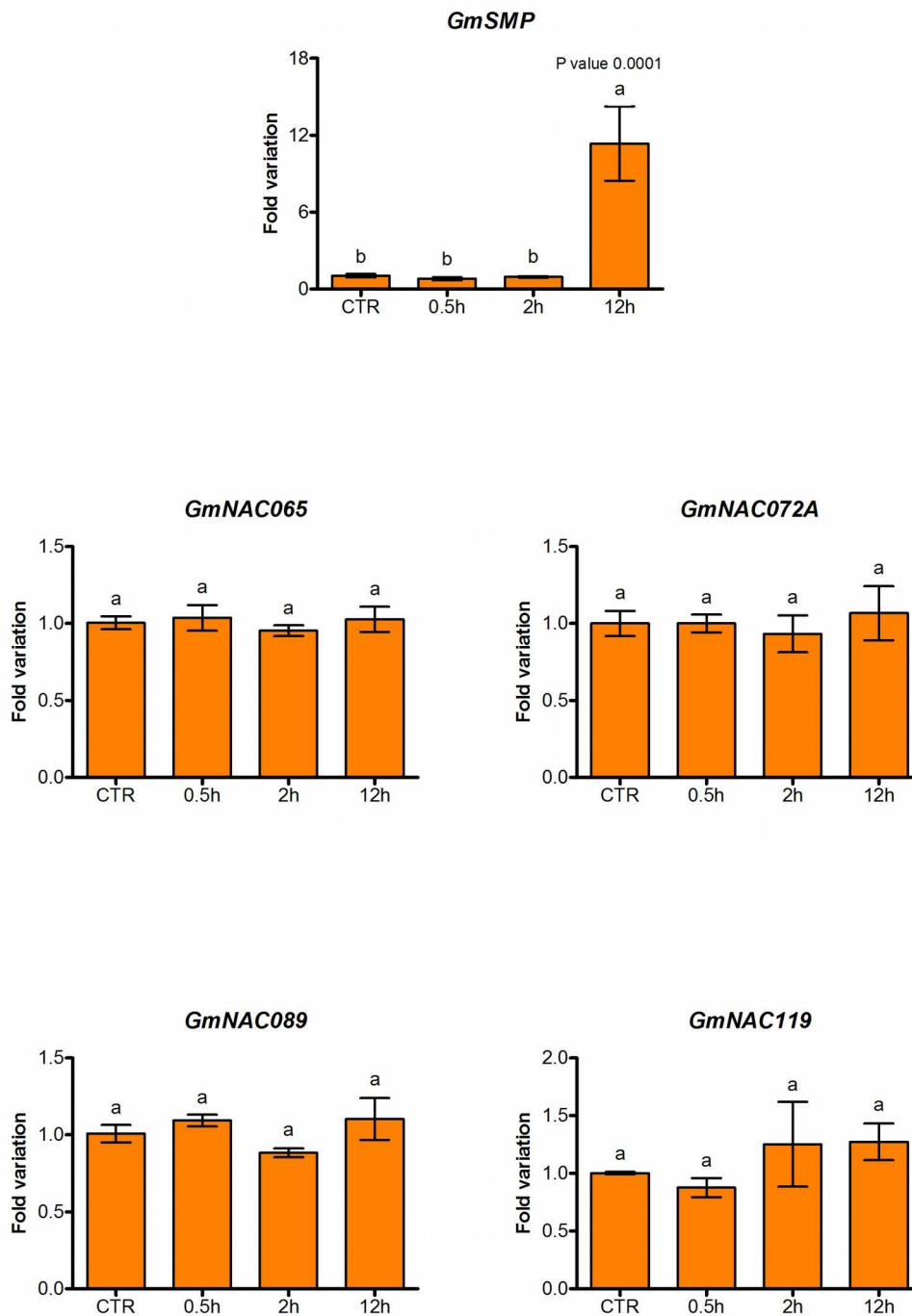


Figura 12. Perfil de expressão gênica sob estresse osmótico simulado por PEG. Plântulas de soja em estágio V3-V4 foram submetidas ao estresse com PEG (PM 8.000 – 0,5 % m/v). No eixo das abcissas encontram-se os tempos de tratamento e a planta não tratada (CTR). No eixo das ordenadas encontra-se a expressão do gene em relação ao CTR. Como teste estatístico entre as amostras biológicas foi realizado o teste de ANOVA one-way com pós teste de Tukey adotando $p < 0,05$. No topo encontra-se a expressão do gene *SMP* (*seed maturation protein*), marcador do estresse osmótico.

representantes são biologicamente diversos e respondem às condições de estresses bastante específicas no âmbito da severidade e o tempo de duração.

NAC089 e NAC119 são proteínas que pertencem a grupos filogenéticos próximos à subfamília NAP e TIP, tradicionalmente envolvidas com o controle da senescência foliar e a resposta contra patógenos, respectivamente [32, 36-37]. A senescência foliar e morte celular programada apresentam uma indução tardia contra estresses bióticos, tais como as infecções virais e fúngicas, além de estresses osmótico e no retículo endoplasmático severos [5, 32-34]. A **Figura 13** mostra o perfil de expressão dos genes frente ao estresse por tunicamicina. Nota-se que tanto *GmNAC089* quanto *GmNAC119* exibem o mesmo perfil de expressão: em plantas estressadas com tunicamicina por 0,5 h há um aumento da expressão dos genes, que diminui em plantas estressadas por 2 h e volta a aumentar (com a maior expressão) em plantas estressadas por 12 h. Esse perfil sugere que esses genes, filogeneticamente relacionados ao controle da senescência foliar, são induzidos com cinética precoce em resposta ao estresse no retículo endoplasmático até que a planta ative a via UPR, com participação de outros fatores, tais como ERD15, as proteínas NRP e a chaperona molecular BiP [5-9]. À medida que o estresse perdura e se torna mais severo, a maquinaria da via UPR é saturada e fatores relacionados ao controle da morte celular programada (cinética tardia da resposta aos estresses) são ativados, tais como NAC081, NAC30 e, possivelmente, NAC089 e NAC119.

GmNAC065, assim como *GmNAC081*, foi filogeneticamente relacionado ao processo de senescência e morte celular programada [34]. Os dados de qRT-PCR mostram que *GmNAC065* e *GmNAC072A* são mais expressos após 12 h de tratamento, quando a severidade do estresse causado por tunicamicina aumenta e pode iniciar a sinalização para vias de senescência e morte celular programada.

Para plantas estressadas com ácido salicílico - que é um hormônio da rota dos fenil-propanoides ativada em resposta aos estresses bióticos - cujo perfil de expressão dos genes se encontra na **Figura 14**, pode-se observar que *GmNAC065* e *GmNAC089* são negativamente regulados à medida que o estresse perdura, diminuindo seu nível de expressão significativamente em plantas estressadas por 12 h. Quando se avalia a resposta das plantas aos estresses bióticos, é difícil definir qual intervalo de tempo melhor define as respostas precoce, intermediária e tardia.

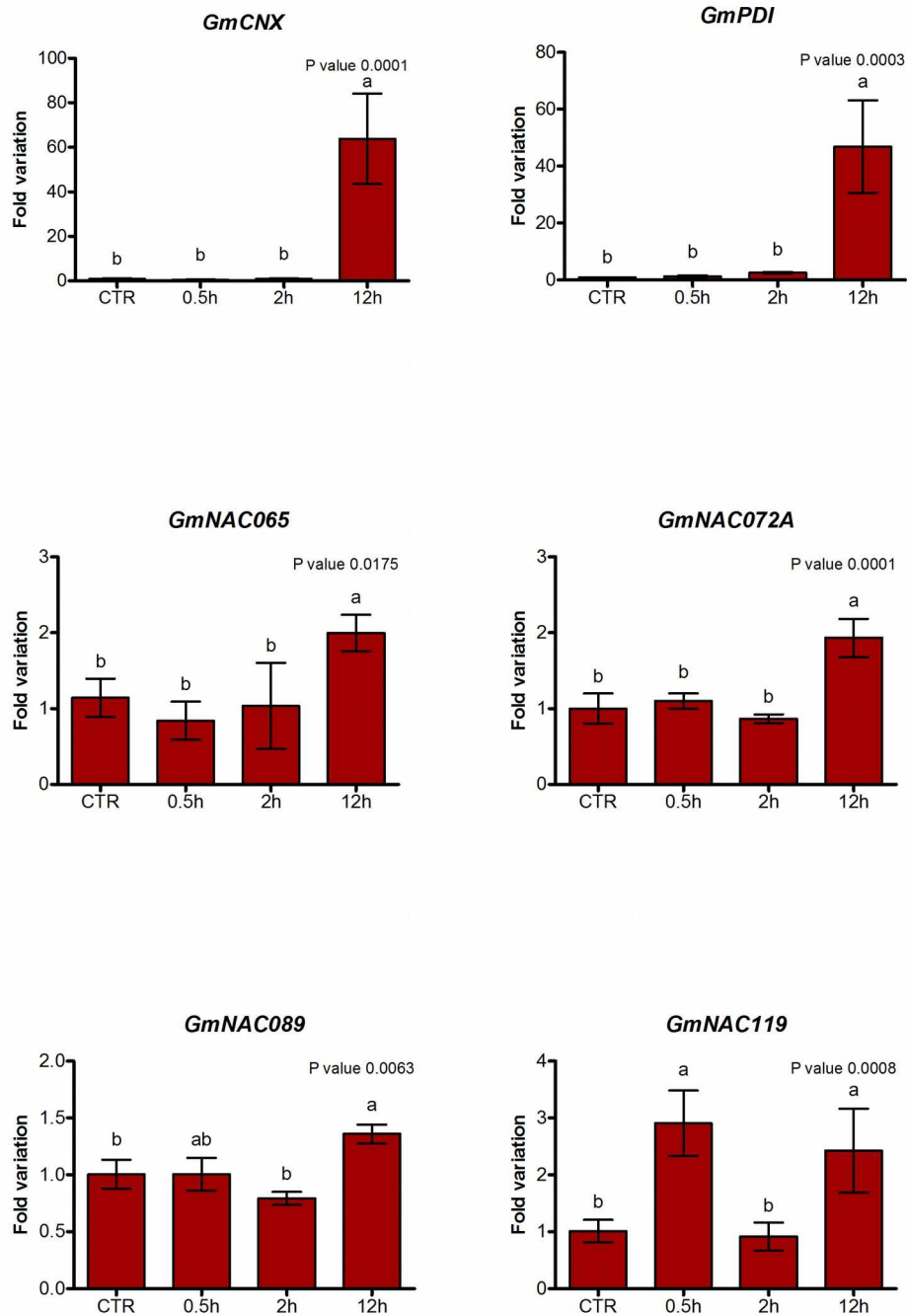


Figura 13. Perfil de expressão gênica sob estresse no retículo endoplasmático simulado por tunicamicina. Plântulas de soja em estágio V3-V4 foram submetidas ao estresse com tunicamicina (5 ug/mL). No eixo das abcissas encontram-se os tempos de tratamento e a planta não tratada (CTR). No eixo das ordenadas encontra-se a expressão do gene em relação ao CTR. Como teste estatístico entre as amostras biológicas foi realizado o teste de ANOVA one-way com pós teste de Tukey adotando $p < 0,05$. Médias significativamente diferentes estão marcadas com as letras a, ab e b. No topo encontra-se a expressão dos genes *CNX* (calnexina) e *PDI* (protein disulfide isomerase), marcadores do estresse no retículo endoplasmático. Os valores de *p value* acima de cada gráfico refere-se ao tratamento com a maior diferença significativa em relação ao controle.

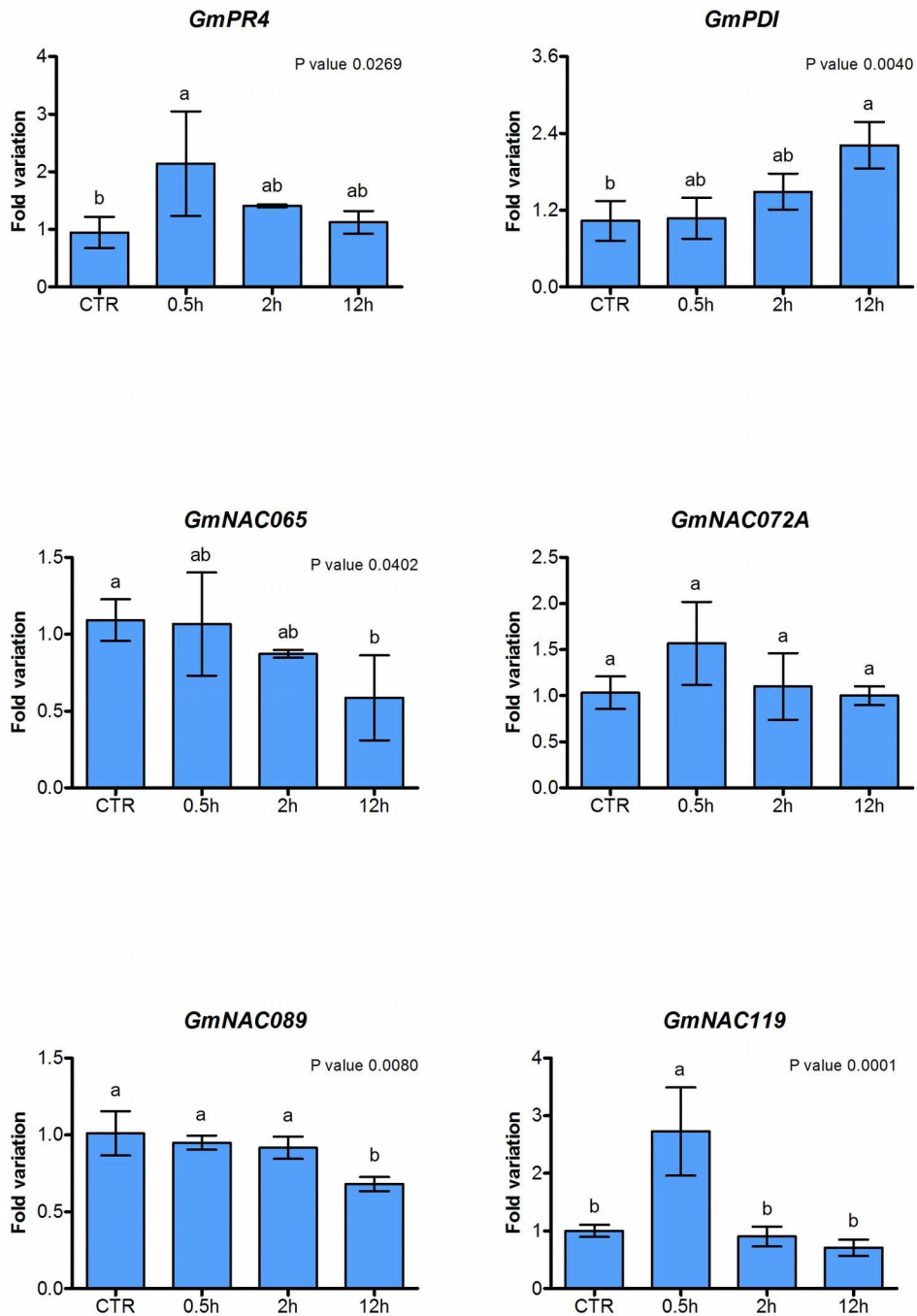


Figura 14. Perfil de expressão gênica sob estresse biótico simulado por ácido salicílico. Plântulas de soja em estágio V3-V4 foram submetidas ao estresse com ácido salicílico (5 mM). No eixo das abcissas encontram-se os tempos de tratamento e a planta não tratada (CTR). No eixo das ordenadas encontra-se a expressão do gene em relação ao CTR. Como teste estatístico entre as amostras biológicas foi realizado o teste de ANOVA one-way com pós teste de Tukey adotando $p < 0,05$. Médias significativamente diferentes estão marcadas com as letras a, ab e b. No topo encontra-se a expressão dos genes *PR-4* (*NPR4*) e *PDI* (*protein disulfide isomerase*), marcadores do estresse biótico mediado por ácido salicílico. Os valores de *p value* acima de cada gráfico refere-se ao tratamento com a maior diferença significativa em relação ao controle.

A interação planta-patogeno é complexa e cada tipo de patossistema envolve diversos fatores que desencadeiam respostas distintas nas plantas na presença de um dado patógeno. Nesse sentido, os perfis de expressão de *GmNAC65* e *GmNAC89* indicam que são genes negativamente regulados com cinética tardia em resposta à sinalização por ácido salicílico. Vale ressaltar que esses genes pertencem as subfamílias ATAF e CUC, respectivamente, que possuem membros relacionados ao controle do desenvolvimento das plantas, como a diferenciação de meristemas [8-9]. Esse perfil é bastante distinto do perfil exibido por *GmNAC072A* e *GmNAC119*. *GmNAC072A* não manifesta qualquer diferença de expressão em nenhum tempo de estresse, levando-nos a supor que não exista o envolvimento entre NAC072 e a resposta a estresse biótico mediada por ácido salicílico. Quanto a *GmNAC119*, percebe-se uma rápida indução em resposta a ácido salicílico, atingindo indução máxima no tempo de 0,5 h e logo declina para padrões normais de expressão a 2 h e 12 h, indicando um envolvimento de NAC119 com os primeiros passos da resposta contra patógenos.

Em uma análise generalizada da expressão dos transfatores em diferentes condições de estresse, nota-se que, para alguns estresses, o perfil de expressão dos genes é antagônico; *GmNAC65*, *GmNAC072A* e *GmNAC089* são induzidos por tunicamicina e reprimidos por ácido salicílico após 12 h de tratamento (à exceção de *GmNAC072A*, cuja expressão não se altera por estes tratamentos). Esses resultados demonstram que os fatores de transcrição da família NAC são expressos em condições específicas em resposta aos estresses em plantas e que sua regulação é importante na elicitação de vias de respostas desencadeadas por diferentes agentes estressantes. Esse antagonismo também é notado na expressão de *GmNAC119*, que no estresse com tunicamicina tem sua expressão aumentada nos tempos alternados de 0,5 h e 12 h de estresse, enquanto que no estresse mediado por ácido salicílico, a expressão do gene só aumenta no tempo de 0,5 h e diminui nos demais tempos de tratamento. Vale ressaltar que os fatores de transcrição NAC072A e NAC089 possuem domínio transmembrana. Essa característica estrutural acrescenta um maior nível na regulação das vias nas quais esses fatores estão envolvidos, uma vez que para que exerçam sua função na reprogramação da expressão gênica eles devem sofrer proteólise ou hidrólise limitadas e deixar o sítio de ancoragem na membrana para o núcleo [35]. Sendo assim, sua atuação se torna dependente da expressão e regulação de proteases

e/ou lipases específicas, inserindo um nível adicional de regulação em relação à atuação de fatores de transcrição solúveis.

Os perfis de expressão de novos genes da superfamília NAC demonstram que os genes, identificados pelas predições *in silico*, devem exercer funções biológicas similares aos membros homólogos com quem se agrupam na árvore filogenética. Nossa análise detectou 32 novos membros, com 5 possíveis pseudo-genes. Dentre os 27 novos genes (verdadeiros), a análise de expressão gênica em diferentes condições de estresse atribuiu uma possível função biológica para 3 novos genes (*GmNAC072A*, *GmNAC089* e *GmNAC119*), correspondendo a cerca de 11% do total de genes novos genes remapeados, além de demonstrar que todos os novos genes examinados são expressos em condições normais. Isso suporta nossa hipótese de que a nova anotação do genoma realmente ofereceu uma atualização à família NAC em soja. O grande número de proteínas codificadas por essa família e a sua grande diversidade permitem às plantas elaborarem respostas complexas, com níveis de regulação hormonal-temporal diversas, e corrobora os dados de análise filogenética, mostrando que as proteínas da família NAC exibem diferentes atividades bioquímicas de acordo com o estágio de desenvolvimento do vegetal e as condições ambientais nas quais ele se encontra.

4. Conclusão

Para o desenvolvimento de ferramentas biotecnológicas e a busca de alvos moleculares que possam ser aplicados na obtenção de cultivares superiores, faz-se necessário entender como as plantas respondem aos estímulos ambientais e como são capazes de alterar sua fisiologia para se adaptarem as situações adversas. Sendo a família NAC um alvo interessante para o entendimento do controle fisiológico das plantas, uma vez que seus genes codificam fatores de transcrição envolvidos com o controle da morfofisiologia do vegetal em condições normais e em condições de estresses abiótico e biótico, procedeu-se a uma análise abrangente da família NAC de soja em dois níveis. Primeiramente, foi feita uma revisão a nova anotação do genoma da soja (versão Wm82.a2.v1) recentemente disponibilizado no Phytozome[®], revelando a presença de 32 possíveis novos membros da superfamília NAC quando comparada ao inventário mais completo dessa família em soja, descrito por Le *et al.* (2011). A família foi atualizada de 152 membros já descritos para 180 membros, sendo 4 genes dentre os 32 encontrados identificados como prováveis repressores de transcrição ou pseudogenes, possuindo apenas o domínio NAC sem domínio de ativação, e 3 novos genes com domínio transmembrana. A superfamília foi filogeneticamente dividida em 5 grandes sub-famílias (ATF, CUC, TIP, TENR e NAP), de acordo com a função de genes ortólogos em *Arabidopsis thaliana* e genes da família já caracterizados em soja. As análises *in silico* mostraram que alguns genes descritos no inventário feito por Le *et al.* (2011) possuíam algum erro de anotação, já que não foram encontrados na varredura feita na nova anotação. Isso levanta a hipótese de que esses genes possam ser pseudogenes, uma vez que não possuem todos os elementos característicos de uma ORF completa. O remapeamento dos genes NAC nos cromossomos da soja revelou que os genes *GmNAC089*, *GmNAC108* e *GmNAC119*, não encontrados em nossa varredura, foram substituídos por novos genes (que receberam os mesmos nomes) e, apenas o gene *GmNAC096* ficou sem seu correspondente na nova anotação.

Em um segundo nível, procedeu-se à validação das análises *in silico* por meio da clonagem de 2 membros já descritos (*GmNAC065*, *GmNAC85*) e um membro novo (*GmNAC119*). Consistente com o papel biológico dos fatores de

transcrição, o transfator GmNAC119, que apresentou nas predições um sinal de localização nuclear, foi localizado no núcleo quando fusionado à GFP e transientemente expresso em folhas de *Nicotiana benthamiana*. NAC119 não demonstrou capacidade de transativação em sistema de duplo híbrido em leveduras, sugerindo que haja em complexo com outras proteínas NAC na ativação de genes alvo. Por fim, o perfil de expressão dos genes *GmNAC65*, *GmNAC072A*, *GmNAC089* e *GmNAC119* foram monitorados em diversas condições cinéticas de respostas aos estresses. Nossos dados mostraram que os genes *GmNAC65*, *GmNAC72A* e *GmNAC119* são positivamente regulados com cinéticas tardias em resposta ao estresse no retículo endoplasmático, enquanto são negativamente regulados também com cinética tardia em resposta aos estresses bióticos mediados por ácido salicílico. Isso remete à grande plasticidade dos fatores de transcrição do tipo NAC, que são regulados em situações fisiológicas específicas da vida das plantas, além de nos fornecer subsídios que validam a revisitação do genoma e o remapeamento dos genes pertencentes à família, atribuindo funções biológicas (preditas pelo agrupamento filogenético) a alguns novos representantes descobertos.

5. Figuras Suplementares

Tabela S1. Membros da superfamília NAC no genoma de soja.

Glyma v11.0	Le et al. 2011	Aminoácidos	Domínio NAC	E-value	Dom. Transm.
Glyma.01G005500.1.p	GmNAC001	451	47 - 189	1.4E-33	
Glyma.01G046800.1.p	GmNAC002	438	16 - 143	4.2E-50	
Glyma.01G051300.1.p	GmNAC003	279	9 - 134	5.3E-52	
Glyma.01G088200.1.p	GmNAC004	451	58 - 197	1.2E-29	
Glyma.01G167900.1.p	GmNAC005	344	9 - 137	9.6E-53	
Glyma.02G050100.1.p	GmNAC006A	362	6 - 135	6.8E-53	
Glyma.02G070000.1.p	GmNAC006	354	9 - 140	2.1E-50	
Glyma.02G070600.1.p	GmNAC007	410	23 - 149	1.8E-53	
Glyma.02G100200.1.p	GmNAC008	436	58 - 197	4.0E-30	
Glyma.02G107000.1.p	GmNAC009	442	18 - 145	4.3E-50	
Glyma.02G109800.1.p	GmNAC010	279	9 134	4.7E-52	
Glyma.13G030900.1.p	GmNAC011	268	9 - 133	7.1E-50	
Glyma.02G222300.1.p	GmNAC012	589	21 - 147	7.9E-50	565 - 587
Glyma.02G240500.1.p	GmNAC013	643	5 - 132	4.9E-54	567 - 586 e 620 - 637
Glyma.02G284300.1.p	GmNAC013A	320	22 - 149	1.0E-47	
Glyma.03G164200.1.p	GmNAC014A	296	10 - 138	6.7E-48	
Glyma.03G179600.1.p	GmNAC014B	287	11 - 138	9.4E-48	
Glyma.03G197900.1.p	GmNAC014	257	4 - 134	1.1E-42	
Glyma.04G014900.1.p	GmNAC015A	350	22 - 149	1.4E-49	
Glyma.04G078600.1.p	GmNAC015	357	28 - 155	2.1E-52	
Glyma.04G119500.1.p	GmNAC016	341	7 - 136	1.6E-53	
Glyma.04G167200.1.p	GmNAC017	357	18 - 143	2.4E-54	
Glyma.04G175800.1.p	GmNAC017A	78	02 -- 78	1.3E-21	
Glyma.04G199000.1.p	GmNAC017B	169	28 - 73	4.0E-11	
Glyma.04G208300.1.p	GmNAC018	291	7 - 131	3.4E-50	
Glyma.04G212000.1.p	GmNAC019	201	8 - 125	6.4E-30	
Glyma.04G213300.1.p	GmNAC020	483	6 - 132	2.9E-46	
Glyma.04G226700.1.p	GmNAC021	603	22 - 148	7.9E-50	579 - 601
Glyma.04G249000.1.p	GmNAC022	300	7 - 131	2.8E-49	
Glyma.05G025500.1.p	GmNAC023	350	16 - 142	3.3E-53	
Glyma.05G055900.1.p	GmNAC024	364	7 - 136	3.4E-53	
Glyma.05G002700.1.p	GmNAC025	390	56 - 197	9.3E-28	
Glyma.05G086000.1.p	GmNAC025A	177	11 - 140	1.5E-48	
Glyma.05G108700.1.p	GmNAC025B	212	18 - 98	6.4E-9	
Glyma.05G113000.1.p	GmNAC026	345	6 - 135	2.0E-50	
Glyma.05G120500.1.p	GmNAC027	189	8 - 127	1.1E-33	
Glyma.05G191300.1.p	GmNAC028	317	5 - 123	1.1E-31	
Glyma.05G192500.1.p	GmNAC029	206	9 - 125	1.7E-30	

(continua na próxima seção)

(Continuação) Tabela S1. Membros da superfamília NAC no genoma de soja.

Glyma.05G195000.1.p	GmNAC030	298	7 - 131	1.3E-49	
Glyma.05G234200.1.p	GmNAC031	321	5 - 131	9.8E-51	
Glyma.05G225100.1.p	GmNAC032	448	47 - 189	9.6E-33	
Glyma.05G202300.1.p	GmNAC033	241	14 - 141	6.2E-44	
Glyma.06G014900.1.p	GmNAC034A	374	22 - 149	1.6E-49	
Glyma.06G080200.1.p	GmNAC034	355	28 - 155	1.8E-52	
Glyma.06G114000.1.p	GmNAC035	299	7 - 131	2.0E-49	
Glyma.06G138100.1.p	GmNAC036	598	22 - 148	6.1E-50	574 - 593
Glyma.06G152900.1.p	GmNAC037	503	6 - 132	7.0E-46	
Glyma.06G154400.1.p	GmNAC038	204	8 - 127	2.3E-31	
Glyma.06G157400.1.p	GmNAC039	295	7 - 131	2.7E-50	
Glyma.06G166500.1.p	GmNAC040	248	14 - 140	4.5E-45	
Glyma.06G195500.1.p	GmNAC041	357	18 - 143	2.0E-54	
Glyma.06G236000.1.p	GmNAC042	375	23 - 151	1.1E-52	
Glyma.06G248900.1.p	GmNAC043	337	14 - 139	2.6E-49	
Glyma.06G249100.1.p	GmNAC044	368	18 - 143	3.8E-51	
Glyma.06G288500.1.p	GmNAC044A	285	49 - 189	3.5E-30	
Glyma.06G318900.1.p	GmNAC045	363	27 - 156	2.5E-53	
Glyma.07G047900.1.p	GmNAC046A	497	12 - 141	2.1E-44	470 - 492
Glyma.07G048000.1.p	GmNAC046	405	6 - 138	3.1E-35	
Glyma.07G048100.1.p	GmNAC047	308	6 - 138	7.0E-37	
Glyma.07G050600.1.p	GmNAC048	400	15 - 143	4.5E-50	
Glyma.07G092000.1.p	GmNAC049	324	5 - 131	4.5E-53	
Glyma.07G126500.1.p	GmNAC050	447	48 - 190	1.1E-32	
Glyma.07G192900.1.p	GmNAC050A	362	15 - 142	3.8E-51	
Glyma.07G201800.1.p	GmNAC051	326	68 - 208	9.7E-31	
Glyma.07G229100.1.p	GmNAC052	233	10 - 135	6.5E-53	
Glyma.07G271100.1.p	GmNAC053	400	43 - 168	2.2E-47	
Glyma.08G009700.1.p	GmNAC054	241	14 - 140	3.7E-44	
Glyma.08G031900.1.p	GmNAC055	452	47 - 189	1.5E-33	
Glyma.08G041500.1.p	GmNAC056	313	5 - 130	1.2E-50	
Glyma.08G075300.1.p	GmNAC057	190	8 - 127	1.3E-33	
Glyma.08G156500.1.p	GmNAC058	323	5 - 130	5.1E-42	
Glyma.08G161300.1.p	GmNAC059	328	10 - 139	6.0E-52	
Glyma.08G163100.1.p	GmNAC059A	348	6 - 133	1.8E-54	
Glyma.08G169400.1.p	GmNAC60	80	27 - 72	1.2E-16	
Glyma.08G173400.1.p	GmNAC61	302	11 - 134	1.1E-49	
Glyma.08G181100.1.p	GmNAC062	190	8 - 128	1.6E-34	
Glyma.08G301100.1.p	GmNAC063	413	15 - 142	3.1E-50	
Glyma.08G307100.1.p	GmNAC064	304	19 - 146	6.8E-48	
Glyma.08G360200.1.p	GmNAC065	224	14 - 139	7.1E-48	

(continua na próxima seção)

(Continuação) Tabela S1. Membros da superfamília NAC no genoma de soja.

Glyma.09G167400.1.p	GmNAC066	265	34 - 161	3.6E-52	
Glyma.09G184100.1.p	GmNAC067	331	5 - 131	3.4E-52	
Glyma.09G231700.1.p	GmNAC068	354	7 - 136	8.7E-54	
Glyma.09G233600.1.p	GmNAC069	359	9 - 138	3.1E-51	
Glyma.09G235700.1.p	GmNAC070	363	19 - 145	1.9E-52	
Glyma.10G037700.1.p	GmNAC071	296	10 - 138	9.3E-48	
Glyma.10G077000.1.p	GmNAC071A	128	11 - 128	1.3E-38	
Glyma.10G077400.1.p	GmNAC071B	136	11 - 136	1.7E-43	
Glyma.10G197500.1.p	GmNAC072	465	33 - 160	8.5E-51	
Glyma.10G197600.1.p	GmNAC072A	448	5 - 97	2.1E-29	424 - 446
Glyma.10G204700.1.p	GmNAC072B	422	56 - 195	1.5E-28	
Glyma.10G216400.1.p	GmNAC073	346	19 - 147	4.2E-51	
Glyma.10G219600.1.p	GmNAC074	560	20 - 146	4.3E-51	529 - 551
Glyma.11G030600.1.p	GmNAC075	360	7 - 138	4.2E-53	
Glyma.11G075400.1.p	GmNAC076	380	40 - 169	9.4E-53	
Glyma.11G096600.1.p	GmNAC077	302	19 - 146	4.5E-45	
Glyma.11G182000.1.p	GmNAC078	246	5 - 137	2.3E-40	
Glyma.11G212400.1.p	GmNAC079	672	8 - 135	1.3E-53	
Glyma.12G003200.1.p	GmNAC080A	356	9 - 138	3.2E-51	
Glyma.12G004900.1.p	GmNAC080	347	7 - 136	1.4E-53	
Glyma.12G022700.1.p	GmNAC081	297	16 - 143	4.3E-45	
Glyma.12G091200.1.p	GmNAC082	248	6 - 140	1.6E-40	
Glyma.12G118700.1.p	GmNAC083	284	49 - 189	6.7E-31	
Glyma.12G145100.1.p	GmNAC083A	180	11 - 138	1.6E-46	
Glyma.12G148900.1.p	GmNAC084	360	14 - 142	2.3E-51	
Glyma.12G149100.1.p	GmNAC085	340	14 - 139	1.7E-49	
Glyma.12G160100.1.p	GmNAC085A	133	2 - 115	5.2E-35	
Glyma.12G161700.1.p	GmNAC086	366	21 - 149	1.1E-52	
Glyma.12G171600.1.p	GmNAC087	349	13 - 139	4.6E-49	
Glyma.12G186200.1.p	GmNAC088	244	4 - 134	2.1E-45	
not found - NAC89					
Glyma.12G186900.1.p	GmNAC089	493	17 - 143	8.8E-47	303 - 325
Glyma.12G206900.1.p	GmNAC090	279	48 - 188	6.1E-30	
Glyma.12G221400.1.p	GmNAC091	375	15 - 142	5.8E-51	
Glyma.12G221500.1.p	GmNAC092	345	14 - 139	4.3E-49	
Glyma.12G226500.1.p	GmNAC093	343	8 - 136	6.7E-53	
Glyma.13G063300.1.p	GmNAC094	380	13 - 138	1.6E-51	
Glyma.13G062000.1.p	GmNAC095	358	20 - 146	6.4E-53	
not found - NAC096					
Glyma.13G174700.1.p	GmNAC097	329	69 - 209	1.7E-30	
Glyma.13G234700.1.p	GmNAC098	332	73 - 214	1.0E-30	

(continua na próxima seção)

(Continuação) Tabela S1. Membros da superfamília NAC no genoma de soja.

Glyma.13G243200.1.p	GmNAC099	363	19 - 147	4.1E-51	
Glyma.13G274300.1.p	GmNAC100	352	16 - 144	7.1E-53	
Glyma.13G279900.1.p	GmNAC101	343	14 - 139	4.2E-49	
Glyma.13G280000.1.p	GmNAC102	375	15 - 142	4.5E-51	
Glyma.13G294000.1.p	GmNAC102A	279	48 - 188	6.1E-30	
Glyma.13G314600.1.p	GmNAC103	371	18 - 144	5.1E-47	331 - 353
Glyma.13G315300.1.p	GmNAC104	253	4 - 134	4.2E-45	
Glyma.13G327600.1.p	GmNAC105	349	13 - 139	4.4E-49	
Glyma.14G030700.1.p	GmNAC106	326	36 - 163	7.7E-48	
Glyma.14G084300.1.p	GmNAC107	278	9 - 142	9.0E-24	
not found - NAC108					
Glyma.14G140100.1.p	GmNAC108	373	27 - 154	4.8E-53	
Glyma.14G152700.1.p	GmNAC109	280	9 - 133	1.1E-49	
Glyma.14G189300.1.p	GmNAC110	590	21 - 147	2.0E-50	566 - 588
Glyma.14G210000.1.p	GmNAC111	644	5 - 132	1.8E-54	568 - 587
Glyma.15G051200.1.p	GmNAC112	191	8 - 128	1.7E-34	
Glyma.15G070300.1.p	GmNAC113	354	16 - 144	1.4E-50	
Glyma.15G078300.1.p	GmNAC114	322	61 - 203	7.9E-31	
Glyma.15G254000.1.p	GmNAC115	303	11 - 134	2.4E-49	
Glyma.15G257700.1.p	GmNAC116	318	27 - 155	4.9E-47	
Glyma.15G264100.1.p	GmNAC117	347	6 - 133	1.6E-54	
Glyma.15G266500.1.p	GmNAC118	326	10 - 139	1.1E-51	
not found - GmNAC119					
Glyma.16G016400.1.p	GmNAC119	267	11 - 140	1.0E-44	
Glyma.16G016600.1.p	GmNAC120	431	5 - 135	7.6E-40	
Glyma.16G016700.1.p	GmNAC121	400	6 - 138	1.1E-36	
Glyma.16G019400.1.p	GmNAC122	443	15 - 143	5.5E-50	
Glyma.16G042900.1.p	GmNAC123	407	16 - 143	2.8E-54	
Glyma.16G043200.1.p	GmNAC124	353	8 - 138	2.7E-51	
Glyma.16G051800.1.p	GmNAC125	216	14 - 140	5.3E-26	
Glyma.16G069300.1.p	GmNAC125A	399	56 - 198	4.8E-27	
Glyma.16G130200.1.p	GmNAC126	362	7 - 135	2.7E-53	
Glyma.16G151500.1.p	GmNAC127	363	9 - 139	1.6E-50	
Glyma.16G152100.1.p	GmNAC128	410	23 - 149	7.1E-54	
Glyma.16G217400.1.p	GmNAC129	264	33 - 160	1.8E-52	
Glyma.17G002800.1.p	GmNAC130	403	47 - 172	2.2E-47	
Glyma.17G101500.1.p	GmNAC131	350	16 - 142	3.3E-53	
Glyma.17G138100.1.p	GmNAC132	366	7 - 136	3.4E-53	
Glyma.17G154100.1.p	GmNAC133	342	6 - 135	3.8E-50	
Glyma.17G185000.1.p	GmNAC134	217	15 - 132	2.0E-31	
Glyma.17G240700.1.p	GmNAC135	285	9 - 143	2.6E-25	

(continua na próxima seção)

(Continuação) Tabela S1. Membros da superfamília NAC no genoma de soja.

Glyma.18G043900.1.p	GmNAC136	678	8 - 135	1.1E-53	580 - 599
Glyma.18G110700.1.p	GmNAC137	304	19 - 146	6.9E-48	
Glyma.18G119300.1.p	GmNAC138	401	15 - 142	4.2E-50	
Glyma.18G261300.1.p	GmNAC139	388	43 - 169	5.3E-51	
Glyma.18G301500.1.p	GmNAC139A	229	14 - 138	1.3E-49	
Glyma.19G002900.1.p	GmNAC139B	389	56 - 197	1.9E-28	
Glyma.19G021900.1.p	GmNAC140	367	14 - 139	1.5E-51	
Glyma.19G024500.1.p	GmNAC141	362	21 - 147	3.0E-53	
Glyma.19G056400.1.p	GmNAC142	398	56 - 198	1.5E-27	
Glyma.19G097700.1.p	GmNAC143	215	14 - 140	6.8E-25	
Glyma.19G108800.1.p	GmNAC143A	336	8 - 138	5.8E-51	
Glyma.19G109100.1.p	GmNAC144	405	16 - 143	2.8E-55	
Glyma.19G165600.1.p	GmNAC144A	294	10 - 138	4.1E-48	
Glyma.19G180300.1.p	GmNAC145	337	13 - 140	1.4E-49	
Glyma.19G195800.1.p	GmNAC145A	254	4 - 134	7.5E-43	
Glyma.19G259500.1.p	GmNAC146	265	16 - 144	7.8E-48	
Glyma.19G259700.1.p	GmNAC147	265	16 - 144	4.8E-48	
Glyma.20G033300.1.p	GmNAC148	280	10 - 135	1.4E-52	
Glyma.20G172100.1.p	GmNAC149	549	20 - 143	2.8E-51	518 - 540
Glyma.20G175500.1.p	GmNAC149A	341	19 - 147	1.6E-51	
Glyma.20G185800.1.p	GmNAC150	442	76 - 215	1.2E-29	
Glyma.20G192300.1.p	GmNAC151	604	5 - 131	1.2E-48	579 - 601
Glyma.20G192500.1.p	GmNAC152	465	34 - 161	8.5E-51	

Legenda de Cores: Amarelo – novos genes candidatos da família NAC identificados em nossa varredura no genoma da soja; Laranja – possíveis pseudogenes; Cinza – proteínas com domínios transmembrana; Verde – genes descritos por Le *et al.* 2011, mas não encontrados na nova varredura do genoma da soja.

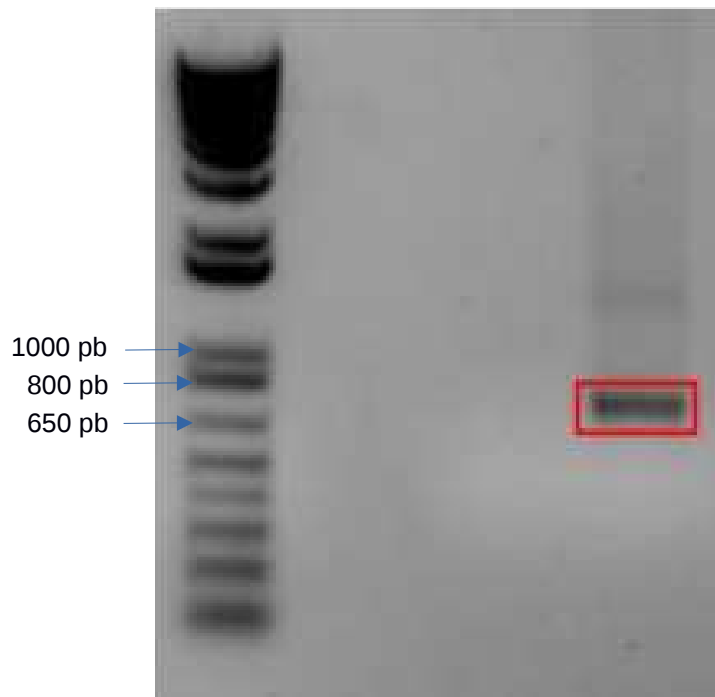


Figura S1. Eletroforese em gel de agarose da amplificação de NAC065. Para amplificação, isolamento e clonagem da ORF de *GmNAC065* (675 pb), duas reações sequenciais de PCR foram realizadas para a amplificação do gene e inserção das regiões de recombinação AttB1/2 (aproximadamente 100 pb) do sistema Gateway[®] de clonagem. À esquerda, o marcador (1 kb ladder) e, à direita, a amplificação após a segunda reação de PCR. Confirmado o tamanho do fragmento amplificado, procedeu-se a purificação do gel de agarose e a reação de recombinação sítio dirigida.

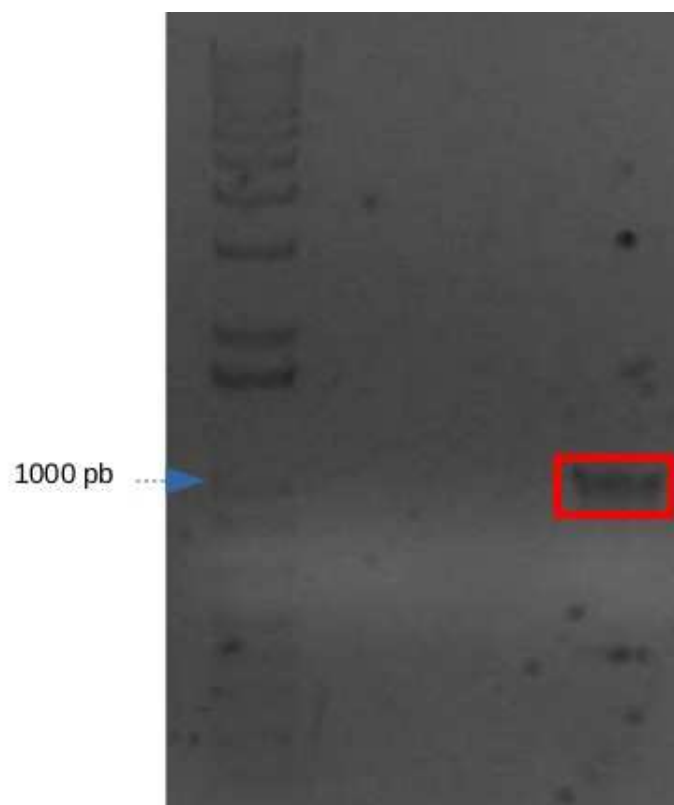


Figura S2. Eletroforese em gel de agarose da amplificação de *GmNAC085*. Para amplificação, isolamento e clonagem da ORF de *GmNAC085* (1023 pb), duas reações sequenciais de PCR foram realizadas para a amplificação do gene e inserção das regiões de recombinação AttB1/2 (aproximadamente 100 pb) do sistema Gateway® de clonagem. À esquerda, o marcador (1 kb ladder) e, à direita, a amplificação após a segunda reação de PCR. Confirmado o tamanho do fragmento amplificado, procedeu-se a purificação do gel de agarose e a reação de recombinação sítio dirigida.

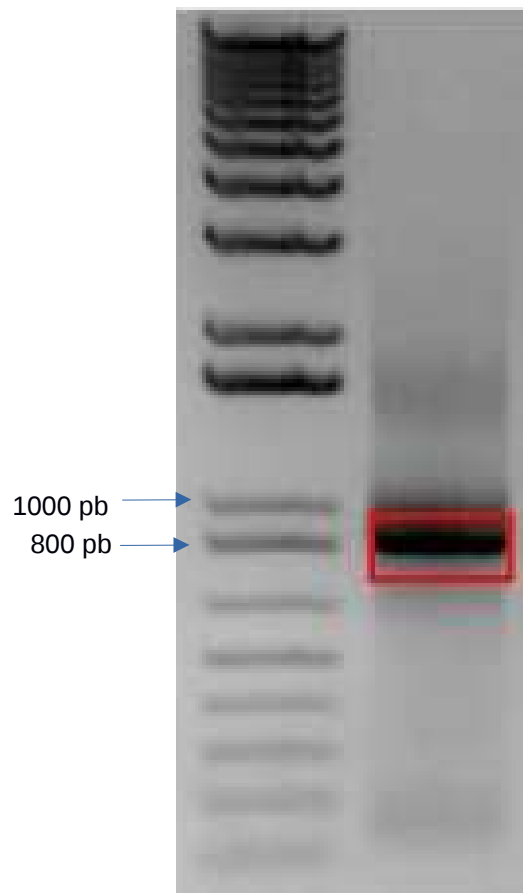


Figura S3. Eletroforese em gel de agarose da amplificação de *GmNAC119*. Para amplificação, isolamento e clonagem da ORF de *GmNAC119* (804 pb), duas reações sequenciais de PCR foram realizadas para a amplificação do gene e inserção das regiões de recombinação AttB1/2 (aproximadamente 100 pb) do sistema Gateway® de clonagem. À esquerda, o marcador (1 kb ladder) e, à direita, a amplificação após a segunda reação de PCR. Confirmado o tamanho do fragmento amplificado, procedeu-se a purificação do gel de agarose e a reação de recombinação sítio dirigida.

```

cDNA_NAC65  ATGGAGAAGGTGAATTTTGTGAAGAATGGAGAGCTTAGATTGCCTCCGGGGTTCCGTTTC
CONTIG      ATGGAGAAGGTGAATTTTGTGAAGAATGGAGAGCTTAGATTGCCTCCGGGGTTCCGTTTC
*****

cDNA_NAC65  CACCCGACTGATGAGGAGCTGGTTCTGCAATACTTGAAGCGCAAGGTCTTCTCCTGCCCT
CONTIG      CACCCGACTGATGAGGAGCTGGTTCTGCAATACTTGAAGCGCAAGGTCTTCTCCTGCCCT
*****

cDNA_NAC65  TTGCCAGCCTCTATCATTCTGAGCTTCATGTTTGCAAGTCTGATCCTTGGGATTTGCCA
CONTIG      TTGCCAGCCTCTATCATTCTGAGCTTCATGTTTGCAAGTCTGATCCTTGGGATTTGCCA
*****

cDNA_NAC65  GGTGATTTGGAGCAAGAGAGATACTTCTTTAGCACCAAAGTGGCCAAATATCCCAACGGA
CONTIG      GGTGATTTGGAGCAAGAGAGATACTTCTTTAGCACCAAAGTGGCCAAATATCCCAACGGA
*****

cDNA_NAC65  AATCGCTCCAACAGAGCCACAAATTCGGGTTATTGGAAGGCAACTGGCTTGGACAAACAA
CONTIG      AATCGCTCCAACAGAGCCACAAATTCGGGTTATTGGAAGGCAACTGGCTTGGACAAACAA
*****

cDNA_NAC65  ATTGTTACTTCAAAGGCAACAACCAAGTTGTCGGAATGAAGAAGACACTTGTTTTCTAC
CONTIG      ATTGTTACTTCAAAGGCAACAACCAAGTTGTCGGAATGAAGAAGACACTTGTTTTCTAC
*****

cDNA_NAC65  AGAGGCAAGCCTCCTAATGGATCCAGAACTGATTGGATCATGCACGAGTATCGCCTCATC
CONTIG      AGAGGCAAGCCTCCTAATGGATCCAGAACTGATTGGATCATGCACGAGTATCGCCTCATC
*****

cDNA_NAC65  CTTAACGCCTCTCAGTCTCAGAGCCATGTTGTTCCCATGGAAAATTGGGTTCTGTGTGCGC
CONTIG      CTTAACGCCTCTCAGTCTCAGAGCCATGTTGTTCCCATGGAAAATTGGGTTCTGTGTGCGC
*****

cDNA_NAC65  ATATTTTTGAAGAGGAGAATTGGTGCTAAAAATGGGGAGGAGAGCAACTCTAAGGTGGTT
CONTIG      ATATTTTTGAAGAGGAGAATTGGTGCTAAAAATGGGGAGGAGAGCAACTCTAAGGTGGTT
*****

cDNA_NAC65  TTCTATGACTTCTAGCGCAGAACAAGACCGATTCTCCTCATCGGTCGCCAGTGGAATT
CONTIG      TTCTATGACTTCTAGCGCAGAACAAGACCGATTCTCCTCATCGGTCGCCAGTGGAATT
*****

cDNA_NAC65  ACACATGAATCAGATGAACATGAAGAGAGCAGTAGCTCCAACACCTTCCCTTATACTATT
CONTIG      ACACATGAATCAGATGAACATGAAGAGAGCAGTAGCTCCAACACCTTCCCTTATACTATT
*****

cDNA_NAC65  AGAAGAAAACCTTAA
CONTIG      AGAAGA-----
*****

```

Figura S4. Confirmação da identidade do clone de *GmNAC065*. O clone gerado foi enviado ao sequenciamento de Sanger para a verificação de sua identidade. O sequenciamento foi realizado utilizando-se cada um dos oligonucleotídeos iniciadores desenhados para a amplificação de cada ORF e a sequência contígua foi alinhada com a sequência da ORF disponível no Phytozome. A ORF de *GmNAC065* foi clonada sem o códon de terminação de tradução.

cDNA_NAC85 CONTIG ATGGGAGTTCCAGAGAGAGACCCTCTTGACAATTGAGCTTGCCCTCCTGGATTTAGATTT
ATGGGAGTTCCAGAGAGAGACCCTCTTGACAATTGAGCTTGCCCTCCTGGATTTAGATTT

cDNA_NAC85 CONTIG TATCCCACTGATGAGGAGCTTTTGGTTTCAGTACCTTTGCCGCAAGGTTGCTGGCCATCAT
TATCCCACTGATGAGGAGCTTTTGGTTTCAGTACCTTTGCCGCAAGGTTGCTGGCCATCAT

cDNA_NAC85 CONTIG TTCTCTCTTCCAATCATTGCTGAAGTTGATTTGTACAAGTTTGATCCATGGGTTCTTCCA
TTCTCTCTTCCAATCATTGCTGAAGTTGATTTGTACAAGTTTGATCCATGGGTTCTTCCA

cDNA_NAC85 CONTIG GGTAAGGCAGCGTTTGGAGAGAAGGAGTGGTACTTCTTCAGTCCAAGAGACAGGAAGTAC
GGTAAGGCAGCGTTTGGAGAGAAGGAGTGGTACTTCTTCAGTCCAAGAGACAGGAAGTAC

cDNA_NAC85 CONTIG CCGAATGGTTCACGACCAAACAGAGTTGCGGGTCTGGGTATTGGAAAGCCACTGGA
ACT CCGAATGGTTCACGACCAAACAGAGTTGCGGGTCTGGGTATTGGAAAGCCACTGGA
ACT *****

cDNA_NAC85 CONTIG GACAAAATCATCACCCTGAAGGTAGAAAAGTTGGCATAAAAAAAGCACTTGTTTTCTAC
GACAAAATCATCACCCTGAAGGTAGAAAAGTTGGCATAAAAAAAGCACTTGTTTTCTAC

cDNA_NAC85 CONTIG GTTGGCAAAGCACCCAAAGGCTCCAAAACCAATTGGATCATGCACGAGTATCGCCTTCTC
GTTGGCAAAGCACCCAAAGGCTCCAAAACCAATTGGATCATGCACGAGTATCGCCTTCTC

cDNA_NAC85 CONTIG GACTCTTCCCGCAAACACAACCTCGGAACCGCAAAGCTTGATGATTGGGTTCTGTGTCGT
GACTCTTCCCGCAAACACAACCTCGGAACCGCAAAGCTTGATGATTGGGTTCTGTGTCGT

cDNA_NAC85 CONTIG ATCTATAAGAAGAACTCAAGTGCGCAAAGGTTGAGGCAAATCTTTTGGCTATGGAATGT
ATCTATAAGAAGAACTCAAGTGCGCAAAGGTTGAGGCAAATCTTTTGGCTATGGAATGT

cDNA_NAC85 CONTIG AGCAATGGGTCATCACCTTCTTCATCGTCCCACGTGGACGACATGCTGGAATCGTTGCCG
AGCAATGGGTCATCACCTTCTTCATCGTCCCACGTGGACGACATGCTGGAATCGTTGCCG

(continua na próxima seção)

```

cDNA_NAC85   GAGATCGATGATCGGTGCTTCACCCTGCCGCGAGTGAAGTCACTCAGTCAGAACAATGCAGCAG
CONTIG       GAGATCGATGATCGGTGCTTCACCCTGCCGCGAGTGAAGTCACTCAGTCAGAACAATGCAGCAG
*****

cDNA_NAC85   CAGGACGAGAAATTCGGATTCAGAACATGGGATCCGGGTTTTTCACCGATTGGGTCAAC
CONTIG       CAGGACGAGAAATTCGGATTCAGAACATGGGATCCGGGTTTTTCACCGATTGGGTCAAC
*****

cDNA_NAC85   CCGACGGATCTTGATTTCAGTTTCCGAATTTGGGTCCGGTTGCCAAACCCAAGGGATGGTG
CONTIG       CCGACGGATCTTGATTTCAGTTTCCGAATTTGGGTCCGGTTGCCAAACCCAAGGGATGGTG
*****

cDNA_NAC85   AATTATGATTGTAATGACTTATTTGTCCCTTCTGTGCCGCCCTTCGGCCACAGCCATGTA
CONTIG       AATTATGATTGTAATGACTTATTTGTCCCTTCTGTGCCGCCCTTCGGCCACAGCCATGTA
*****

cDNA_NAC85   AACTACATGGTGGGGGCACCACCGTCCGAGGAGGAGTTCAAAGCGGTGTGAGGACTCAA
CONTIG       AACTACATGGTGGGGGCACCACCGTCCGAGGAGGAGTTCAAAGCGGTGTGAGGACTCAA
*****

cDNA_NAC85   CAGGCCGATGGGGCCGCATGTTTTTCAGCAGAACCCAAATGCCCGATTGTTACCGGGCTCG
CONTIG       CAGGCCGATGGGGCCGCATGTTTTTCAGCAGAACCCAAATGCCCGATTGTTACCGGGCTCG
*****

cDNA_NAC85   GGCGACCCATTTGGGTTTGGGTTTCATCATGGGTCAGCAAGTTGAGTTCGGGTTTAGGGAC
CONTIG       GGCGACCCATTTGGGTTTGGGTTTCATCATGGGTCAGCAAGTTGAGTTCGGGTTTAGGGAC
*****

cDNA_NAC85   TGA -----
CONTIG       TGA -----
***

```

Figura S5. Confirmação da identidade do clone de *GmNAC085*. O clone gerado foi enviado ao sequenciamento de Sanger para a verificação de sua identidade. O sequenciamento foi realizado utilizando-se cada um dos oligonucleotídeos iniciadores desenhados para a amplificação de cada ORF e a sequência contígua foi alinhada com a sequência da ORF disponível no Phytozome.

cDNA_NAC119 CONTIG ATGGAGAACATAATTTCCATATGTGATCACATGCCTGTAGGATTCAGATTCCGTCCCACA
ATGGAGAACATAATTTCCATATGTGATCACATGCCTGTAGGATTCAGATTCCGTCCCACA

cDNA_NAC119 CONTIG GATGAAGAGCTTGTCGACTATTACCTCAAACACAAGCTTCTGGCTGATGATTTCCCTGTC
GATGAAGAGCTTGTCGACTATTACCTCAAACACAAGCTTCTGGCTGATGATTTCCCTGTC

cDNA_NAC119 CONTIG CACATCATCCCTGAAATCGACCTTTGCAAGGTGGAACCTTGGGACGTACCAGGAAGATCA
CACATCATCCCTGAAATCGACCTTTGCAAGGTGGAACCTTGGGACGTACCAGGAAGATCA

cDNA_NAC119 CONTIG GTGATCAAATCCGATGACCCAGAATGGTTTTTCTTCAGTCCTGTGGATTACAAGTACTTG
GTGATCAAATCCGATGACCCAGAATGGTTTTTCTTCAGTCCTGTGGATTACAAGTACTTG

cDNA_NAC119 CONTIG AAGAGTAAAAGGTTTAAACAGGACAACCAAGCGTGGCTTCTGGAAAGCCACTGGGAATGAC
AAGAGTAAAAGGTTTAAACAGGACAACCAAGCGTGGCTTCTGGAAAGCCACTGGGAATGAC

cDNA_NAC119 CONTIG CGGAAAATAAGGATACCGGGCACCAGCAATGTCATTGGGACTAAGAAGACTCTTGTTTTTC
CGGAAAATAAGGATACCGGGCACCAGCAATGTCATTGGGACTAAGAAGACTCTTGTTTTTC

cDNA_NAC119 CONTIG CACCAAGGCCGTGTTCCCCGTGGTGCCAAGACCAACTGGGTTATTCATGAATACCATGCT
CACCAAGGCCGTGTTCCCCGTGGTGCCAAGACCAACTGGGTTATTCATGAATACCATGCT

cDNA_NAC119 CONTIG GTTACCTCCCATGAAAGCCAGAGGGCTTTTTGTTTTGTGCCGCTTGATGAAGAAAGCTGAG
GTTACCTCCCATGAAAGCCAGAGGGCTTTTTGTTTTGTGCCGCTTGATGAAGAAAGCTGAG

cDNA_NAC119 CONTIG AAAAAAATGAAGGGGGAATTGAGGCACCGTCTTTTGATGAAGGGGAGCCCAGTATACAC
AAAAAATGAAGGGGGAATTGAGGCACCGTCTTTTGATGAAGGGGAGCCCAGTATACAC

cDNA_NAC119 CONTIG ATGGTTTCTGACTATGGAAATCCGGCAAGAGCAGAAGGCATTGCACCTGTAGATATTATT
ATGGTTTCTGACTATGGAAATCCGGCAAGAGCAGAAGGCATTGCACCTGTAGATATTATT

(continua na próxima seção)

```

cDNA_NAC119    CCTGGAGTAGATATGGATGAAATCTTTCGGGCAGTGGATCAAGCTGAGAAATTCTCCCC
CONTIG         CCTGGAGTAGATATGGATGAAATCTTTCGGGCAGTGGATCAAGCTGAGAAATTCTCCCC
                *****

cDNA_NAC119    CCAGTACAACAGTCTCCAACCTGGCATTGGGCTGGAAGATTTTTTCCCAAACCTCTCCATTG
CONTIG         CCAGTACAACAGTCTCCAACCTGGCATTGGGCTGGAAGATTTTTTCCCAAACCTCTCCATTG
                *****

cDNA_NAC119    TTCAATGCCCGTTTTTGAAGTGAAAACATCAATATGCAAACCTCACTTGAAGTTACGGAT
CONTIG         TTCAATGCCCGTTTTTGAAGTGAAAACATCAATATGCAAACCTCACTTGAAGTTACGGAT
                *****

cDNA_NAC119    GACGAAGATGAGTTCATACCCTAG-----
CONTIG         GACGAAGATGAGTTCATACCC-----
                *****

```

Figura S6. Confirmação da identidade do clone de *GmNAC119*. O clone gerado foi enviado ao sequenciamento de Sanger para a verificação de sua identidade. O sequenciamento foi realizado utilizando-se cada um dos oligonucleotídeos iniciadores desenhados para a amplificação de cada ORF e a sequência contígua foi alinhada com a sequência da ORF disponível no Phytozome. A ORF de *GmNAC118.A* foi clonada sem o códon de terminação de tradução.

6. Referências Bibliográficas

- [1] **Tran, L.S. and Mochida, K.** (2010). Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean. *Plant signaling and behavior* 5 – 255-257.
- [2] **Le, D.T.; Nishiyama, R.; Watanabe, Y.; Mochida, K.; Yamaguchi-Shinozaki, K.; Shinozaki, K.; Tran, L.S.** (2001). Genome wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Res* 18 – 263-276.
- [3] **Olsen, A.N.; Ernst, H.A.; Leggio, L. L.; Skriver, K.** (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science* 10 – 79-87.
- [4] **Nakashima, K.; Takahashi, H.; Mizoi, J.; Shinozaki, K.; Yamaguchi-Shinozaki, K.** (2012). NAC transcription factors in plant abiotic stress response. *Biochimica et biophysica acta* 1818 – 97-103.
- [5] **Mendes, G.C.; Reis, P.A.B.; Calil, I.P.; Carvalho, H.H.; Aragão, F.J.L.; Fontes, E.P.B.** (2013). GmNAC 30 and GmNAC81 integrate the endoplasmatic reticulum stress-and osmotic stress-induced cell death responses through a vacuolar processing enzyme. *PNAS* 110/48 19627-19632.
- [6] **Zhu, J.K.** (2002). Salt and drought signal transduction in plants. *Annual Review of Plant Physiology. Plant Molecular Biology* - 53 - 247 – 273.
- [7] **Souer, E.; van Houwelingen, A.; Kloss, D.; Mol, J.; Koes, R.** (1996). The non apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* 85 - 159 – 170.
- [8] **Aida, M.; Ishida, T.; Fukaki, H.; Fujisawa, H.; Tasaka, M.** (1997). Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. *Plant Cell* 9 - 841 – 857.

- [9] **Sablowski, R.W.M., Meyerowitz, E.M.** (1998). A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PITILLATA. *Cell* 92 - 93 – 103.
- [10] **Xie, Q.; Frugis, G.; Colgan, D.; Chua, N.H.** (2000). *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev.* 14 - 3024 – 3036.
- [11] **Reis, P.A. and Fontes, E.P.B.** (2012). N-rich protein (NRP)-mediated cell death signaling: a new branch of the ER stress response with implications for plant biotechnology. *Plant signaling and behavior* 7 – 628-632.
- [12] **Alves, M.S.; Reis, P.A.B.; Dadalto, S. P.; Faria, J.A.Q.A.; Fontes, E.P.B.F.; Fietto, L.G.** (2011). A novel transcription factor, ERD15 (early responsive to dehydration 15), connects endoplasmic reticulum stress with an osmotic stress-induced cell death signal. *The journal of biological chemistry* 286, 20020 – 20030;
- [13] **Reis, P.A.; Rosado, G.L.; Silva, L.A.; Oliveira, L.B.; Costa, M.D.; Alvim, F.C.; Fontes, E.P.** (2011). The binding protein BiP attenuates stress-induced cell death in soybean via modulation of the N-rich protein-mediated signaling pathway. *Plant Physiology* 157, 1853 – 1865.
- [14] **Carvalho, H.H.; Brustolini, O.J.B.; Pimenta, M.R.P.; Mendes, G.C.M.; Gouveia, B.C.; Silva, P.A.; Silva, J.C.F.; Mora, C.S.; Soares-Ramos, J.R.L.; Fontes, E.P.B.** (2014). Molecular chaperone binding protein BiP prevents leaf dehydration-induced cellular homeostasis disruption. *Plos One* 9 - 1 – 18.
- [15] **Tran, L.S.; Nakashima, K.; Sakuma, Y.; Simpson, S.D.; Fujita, Y.; Maruyama, K.; Fujita, M.; Seki, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K.** (2004). Isolation and functional analysis of *Arabidopsis* stress-induced NAC transcription factors that bind to a drought-responsive cis element in the EARLY RESPONSIVE TO DEHYDRATION STRESS 1 promoter. *Plant Cell* 16 - 2481 – 2498.

- [16] **Fujita, M.; Fujita, Y.; Maruyama, K.; Seki, M.; Hiratsu, K.; Ohme-Takagi, M.; Tran, L.S.; Yamaguchi-Shinozaki, K.; Shinozaki, K.** (2004). A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant Journal* 39 - 863 – 876.
- [17] **Nakashima, K.; Ito, Y.; Yamaguchi-Shinozaki, K.** (2009). Transcription regulatory network in response to abiotic stresses in *Arabidopsis* and Grasses. *Plant Physiology* 149 – 88 – 95.
- [18] **Jeong, J.S.; Kim, Y.S.; Baek, K.H.; Jung, K.; Ha, S.H.; Do Choi, Y.; Kim, M.; Reuzeau, C.; Kim, J.K.** (2010). Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiology* 153 – 185 – 197.
- [19] **Hu, H.; Dai, M.; Yao, J.; Xiao, B.; Li, X.; Zhang, Q.; Xiong, L.** (2006). Over-expressing a NAM, ATAF and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *PNAS* 103/35 – 12987 – 12992.
- [20] **Udupa, S. M.; Robertson, L.D.; Weigand, F.; Baum, M.; Kahl, G.** (1999). Allelic variation at (TAA)_n microsatellite loci in a world collection of chickpea germoplasm. *Molecular and General Genetic* 261 – 354 – 363.
- [21] **Fang, Y.; Liao, K.; Du, H.; Xu, Y.; Song, H.; Li, X.; Xiong, L.** (2015). A stress-responsive transcription factor SNAC3 confers heat and drought tolerance through modulation of reactive oxygen species in rice. *Journal of Experimental Botany* 66/21 – 6803 – 6817.
- [22] **Lee, S.; Seo, P.J.; Lee, H.J.; Park, C.M.** (2012). A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in *Arabidopsis*. *Plant Journal* 70/5 – 831 – 844.
- [23] **Wu, A.; Allu, A.D.; Garapati, P.; Siddiqui, H.; Dortay, H.; Zanol, M.I.; Asensi-Fabado, M.A.; Munné-Bosh, M.A.; Antonio, C.; Tohge, T.; Fernie, A.R.;**

Kaufmann, K.; Xue, G.P.; Mueller-Roeber, B.; Balazadeh, S. (2012). JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. *Plant Cell* 24/2 – 482 – 506.

[24] **Kariola, T.; Brader, G.; Helenius, E.; Li, J.; Heino, P.; Palva, E. T. (2006).** Early Responsive to Dehydration 15, a Negative Regulator of Abscisic Acid Responses in Arabidopsis. *Plant Physiology* 142 – 1559 – 1573.

[25] **Pascual, M. B.; Cánovas, F. M.; Ávila, C. (2015).** The NAC transcription factor family in maritime pine (*Pinus Pinaster*): molecular regulation of two genes involved in stress responses. *BMC Plant Biology* 15 – 254.

[26] **Hao, Yu-Jun; Wei, W.; Song, Q.X.; Chen, W.H.; Zhang, Y.Q.; Wang, F.; Zou, H.F.; Lei, G.; Tian, A.G.; Zhang, W.K.; Ma, B.; Zhang, J.S.; Chen, S.Y. (2011).** Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. *The Plant Journal* 68 - 302 – 313.

[27] **Nuruzzaman, M.; Sharoni, A.M.; Satoh, K.; Moumeni, A.; Venuprasad, R.; Serraj, R.; Kumar, A.; Leung, H.; Attia, K.; Kikuchi, S. (2012).** Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64. *Molecular Genetics and Genomics* 287/5 - 389 – 410.

[28] **Takasaki, H.; Maruyama, K.; Takahashi, F.; Fujita, M.; Yoshida, T.; Nakashima, K.; Myouga, F.; Toyooka, K.; Yamaguchi-Shinozaki, K.; Shinozaki, K. (2015).** SNAC-As, stress-responsive NAC transcription factors, mediate ABA-inducible leaf senescence. *The Plant Journal* 84 – 1114 – 1123.

[29] **Delessert, C.; Kazan, K.; Wilson, I.W.; Van der Straeten, D. Manners, J.; Dennis, E.S.; Dolferus, R. (2005).** The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis. *The Plant Journal* 43/5 – 745 – 757.

- [30] **Jensen, M. K.; Rung, J.H.; Gregersen, P.L.; Gjetting, T.; Fuglsang, A.T.; Hansen, M.; Joehnk, N.; Lyngkjaer, M.F.; Collinge, D.B.** (2007). The HvNAC6 transcription factor: a positive regulator of penetration resistance in barley and *Arabidopsis*. *Plant Molecular Biology* 65 – 137 – 150.
- [31] **Kim, H. J.; Nam, H.G.; Lim, P.O.** (2016). Regulatory network of NAC transcription factors in leaf senescence. *Current opinion in plant biology* 33 – 48 – 56.
- [32] **Pinheiro, G. L.; Marques, C.; Costa, M.D.L.; Reis, P.A.B.; Alves, M.S.; Carvalho, C.; Fietto, L.G.; Fontes, E.P.B.** (2009). Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene* 444 – 10 – 23.
- [33] **Faria, J. A. Q. A.; Reis, P.A.B.; Reis, M.T.B.; Rosado, G.L.; Pinheiro, G.L.; Mendes, G.C.; Fontes, E.P.B.** (2001). 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 Biology* 11 – 129.
- [34] **Pimenta, M. R.; Silva, P.A.; Mendes, G.C.; Alves, J.R.; Caetano, H. D.; Machado, J.P.B.; Brustolini, O.J.; Carpinetti, P.A.; Melo, B.P; Silva, J.C; Rosado, G.L; Ferreira, M.F.; Dal-Bianco, M.; Picoli, E.A.; Aragão, F.J.; Ramos, H.J.; Fontes, E.P.B.** (2016). The Stress-Induced Soybean NAC Transcription Factor GmNAC81 Plays a Positive Role in Developmentally Programmed Leaf Senescence. *Plant and Cell physiology* 57 – 1098 – 1114.
- [35] **Seo, P. J.; Kim, M.J.; Song, J.S.; Kim, Y.S.; Kim, H.J.; Park, C.M.** (2010). Proteolytic processing of an Arabidopsis membrane-bound NAC transcription factor is triggered by cold-induced changes in membrane fluidity. *Biochemical Journal* 427/3 – 359 – 367.
- [36] **Wang, X.; Basnayake, B.M.; Zhang, H.; Li, G.; Li, W.; Virk, N.; Mengiste, T.; Song, F.** (2009). The Arabidopsis ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens. *Molecular Plant-Microbe Interaction* 22 – 1227 – 1238.

[37] **Guo, Y. and Gan, S.** (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *The Plant Journal* 46/4 – 601 – 612.

[38] **Yamaguchi, Ohtani, M.; Mitsuda, N.; Kubo, M.; Ohme-Takagi, M.; Fukuda, H.; Demura, T.** (2010). VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in *Arabidopsis*. *Plant Cell* 22/4 – 1249 – 63.