

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

RÔMULO DA FONSÊCA DOS SANTOS

**ANÁLISE ESTRUTURAL E FUNCIONAL DE FATORES DE TRANSCRIÇÃO DA
FAMÍLIA NAC EM FEIJÃO-CAUPI (VIGNA UNGUICULATA) E OUTRAS
LEGUMINOSAS**

Recife

2017

RÔMULO DA FONSÊCA DOS SANTOS

**ANÁLISE ESTRUTURAL E FUNCIONAL DE FATORES DE TRANSCRIÇÃO DA
FAMÍLIA NAC EM FEIJÃO-CAUPI (VIGNA UNGUICULATA) E OUTRAS
LEGUMINOSAS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Mestre em Ciências Biológicas.

Orientadora: Dr^a. Ana Maria Benko Iseppon

Coorientadora: Dr^a. Valesca Pandolfi

Recife

2017

Catálogo na fonte:

Bibliotecário Bruno Márcio Gouveia - CRB-4/1788

Santos, Rômulo da Fonseca

Análise estrutural e funcional de fatores de transcrição da família NAC em feijão-caupi (*Vigna unguiculata*) e outras leguminosas / Rômulo da Fonseca Santos. – 2017.

86 f. : il.

Orientadora: Prof^a. Dr^a. Ana Maria Benko Iseppon.

Coorientadora: Prof^a. Dr^a Valesca Pandolfi.

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Ciências Biológicas, Recife, 2017. Inclui referências.

1. Leguminosa. 2. Feijão-caupi. 3. Genética vegetal. I. Iseppon, Ana Maria Benko (Orientadora). II. Pandolfi, Valesca (Coorientadora). III. Título.

683.74

CDD (22.ed.)

UFPE/CB – 2019 - 161

RÔMULO DA FONSÊCA DOS SANTOS

**ANÁLISE ESTRUTURAL E FUNCIONAL DE FATORES DE TRANSCRIÇÃO DA
FAMÍLIA NAC EM FEIJÃO-CAUPI (VIGNA UNGUICULATA) E OUTRAS
LEGUMINOSAS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Mestre em Ciências Biológicas.

Aprovado em: 23/02/2017

BANCA EXAMINADORA

Dr^a. Ana Maria Benko Iseppon (Orientadora)
Universidade Federal de Pernambuco

Dr. José Ribamar Costa Ferreira Neto
Universidade Federal de Pernambuco

Dr. João Pacífico Bezerra Neto
Universidade Federal de Pernambuco

À minha família, por todo amor dedicado, sacrifícios realizados,
ensinamentos compartilhados e vitórias alcançadas. Por serem
meu porto seguro em meio às tempestades.
Dedico.

AGRADECIMENTOS

Agradeço em primeiro lugar a **Deus**, nosso pai eterno e autor da minha vida, cujo amor incondicional e misericórdia sem fim permitiu que, mesmo com todas as batalhas, alcançasse mais essa vitória.

Agradeço a minha mãe, **Daída Maria da Fonsêca dos Santos**, por todo o amor dedicado a mim e por todos seus ensinamentos. Agradeço por ter acreditado em mim quando ninguém mais acreditou. Agradeço por sempre segurar na minha mão e me ajudar incondicionalmente na conquista dos meus objetivos. Sem o seu exemplo nada disso seria possível.

Ao meu pai, **Normando Augusto dos Santos** (*in memoriam*), por todo o sacrifício realizado em vida para que os meus sonhos se tornassem possíveis e pelo exemplo que sempre foi.

A minha irmã, **Rafaela da Fonsêca dos Santos**, que embora mais nova, sempre me apoio e me incentivou inclusive nas horas mais difíceis. Por todo o seu amor e companhia. Por ter sido a minha companheira de aventuras e ter feito do inevitável desafio de crescer, um fardo mais leve.

Ao meu avô, **Alcides Augusto dos Santos** (*in memoriam*), meu primeiro e único herói. Pela inspiração que sempre foi e ainda é em minha vida. Pelas longas conversas e ensinamentos. Pelo apoio incondicional e sacrifícios realizados para que nada me faltasse. Por ter me ensinado o valor e a importância de estudar e adquirir conhecimentos para a vida, e não para provas. Agradeço por “nunca me deixar num mato sem cachorro”.

As minhas avós, **Maria Queiroz dos Santos** (*in memoriam*) e **Juracy da Fonsêca Pereira** (*in memoriam*). Sem o amor e carinho de vocês, tudo teria sido mais difícil. Por terem sido verdadeiros tesouros imateriais em minha vida. Com vocês pude aprender a sempre me dedicar para aliviar o fardo do próximo quando este for pesado.

A minha tia **Nerise**, por ser uma segunda mãe. Mesmo nos momentos mais difíceis e, apesar da distância, seu amor e dedicação foram essenciais para mais essa vitória.

Aos meus familiares, em especial aos meus primos/irmãos **Paulo de Tarso** e **Paulo Filho**. Pela inspiração e exemplo que sempre foram para mim. Graças ao convívio com vocês, o aprendizado de inúmeras coisas se fez possível.

Agradeço enormemente a Prof^a Dr^a **Ana Maria Benko Iseppon**, por sua orientação e pelo exemplo de pessoa e profissional. Pelos conselhos e críticas, ensinamentos e oportunidades. Sem sua dedicação, o meu crescimento acadêmico teria sido mais difícil.

Agradeço em especial a Dr^a **Valesca Pandolfi**, a minha “mãe científica” e amiga verdadeira. Por sua coorientação e por todos os momentos dentro e fora do laboratório. Pelo amor incondicional e por todas as conversas e conselhos pessoais e profissionais. Por ter acreditado e não ter desistido de mim. Por ter sido minha companheira mesmo nas horas mais difíceis. Sem o seu apoio essa conquista não teria sido possível.

Ao Dr. **João Pacífico**, por todos os ensinamentos desde os tempos de Iniciação Científica. Pelo exemplo de pessoa e profissional que sempre foi. Pelo socorro sempre presente e por toda a dedicação.

Aos meus colegas do LGBV, em especial **Marx Lima, Flávia Araújo e Roberta Lane** (a velha guarda), pelos ensinamentos e momentos de descontração. Por serem mais que colegas de trabalho e por transformarem o meu dia-a-dia no laboratório mais leves. Espero um dia poder retribuir tudo o que sempre fizeram por mim.

Aos professores das Ciências Biológicas da Universidade Federal de Pernambuco. Em especial ao Dr. **Bruno Severo** pelo exemplo de pessoa e profissional que sempre foi para mim e muitos outros alunos. Ao Dr. **Otacílio** (meu “irmão”), pela inspiração e por ter contribuído significativamente na minha formação como biólogo.

Aos meus colegas biólogos e companheiros de jornada, **Arthur Cabral, Leonardo Bento, Júlio Cesar, Fábio Gomes, Fausto Pedro, Luana Elizabett, Thiago Oliveira, Guilherme Pereira e Laura Bento**, por sua amizade e companhia, nas horas boas e ruins, durante toda a graduação e mestrado. Vocês ajudaram a tornar essa caminhada mais fácil. Em especial, a **Nathaly Santos**, por sua amizade verdadeira, incentivo e confiança em mim depositada. Agradeço a Deus pela vida e amizade de cada um de vocês.

Aos missionários de A Igreja de Jesus Cristo dos Santos dos Últimos Dias, **Elders Valdez, dos Santos, H. Felipe, F. Santos, Gomes (Jorge), Sanhueza e Jex**. A amizade, o amor e o exemplo de vocês foram determinantes para o meu crescimento pessoal, profissional e espiritual, mesmo nas horas de descontração. Agradeço em especial aos **Elders Barbella e G. Azevedo**, por terem batido à minha porta naquela tarde de 26 de agosto de 2015. Foi através de vocês que pude voltar ao caminho do bem e da verdade.

Ao suporte financeiro da agência de fomento à pesquisa CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), no âmbito do edital CAPES/Embrapa nº. 15/2014.

E a todos aqueles que de alguma forma contribuíram para a realização e conclusão deste trabalho.

Muito obrigado.

“E assim chego a uma concepção de relação entre ciência e
religião muito diferente da usual... Sustento que o
sentimento religioso cósmico é a mais forte motivação da
pesquisa científica.”

Albert Einstein

RESUMO

Genes da família NAC (NAM, ATAF1/2 e CUC2) codificam fatores de transcrição (FTs) específicos de plantas, com importante papel na resposta a estresses bióticos quanto abióticos, bem como no crescimento e desenvolvimento vegetal. Neste trabalho, sequências NAC de *Arabidopsis thaliana* e de quatro leguminosas (Fabaceae) foram utilizadas para a identificação e caracterização estrutural e funcional de FTs-NAC no transcriptoma da soja e do feijão-caupi. Foram identificadas 50 e 42 FTs NAC nos transcriptomas da soja e do feijão-caupi, respectivamente, submetidos a desidratação radicular, comparativamente aos controles não estressados. As proteínas correspondentes aos genes candidatos foram avaliadas quanto à presença de domínios conservados, sendo classificadas em cinco grupos (NAC-*a* até NAC-*e*), sendo NAC-*d* o grupo mais representativo, enquanto NAC-*c* apresentou apenas cinco sequências de feijão-caupi e nenhuma de soja. As proteínas apresentaram relativa conservação especialmente na região 5' ao lado de maior variação na região 3', refletindo na estrutura secundária revelada pela modelagem tridimensional aplicada a nove proteínas (cinco de feijão-caupi e quatro de soja). A análise da expressão dos FTs NAC revelou candidatos induzidos em todos tecidos e situações (inclusive controles não estressados), enquanto outros apresentaram baixa expressão ou especificidade de tempo ou tecido. Os resultados indicam significativa diversidade desses fatores e diferentes estratégias nas leguminosas analisadas, sendo potencialmente úteis para uso como marcadores moleculares ou em transgênese, visando melhor desempenho sob estresses bióticos e abióticos.

Palavras-chave: Estresse abiótico. Feijão-caupi. Soja (*Glycine max*). Desidratação radicular. Bioinformática.

ABSTRACT

The NAC gene family (NAM, ATAF1 / 2 and CUC2) encodes plant-specific transcription factors (FTs) with an important role in the response to biotic and abiotic stresses as well as in growth and plant development. In this work, NAC seed sequences from *Arabidopsis thaliana* and four leguminous plants (Fabaceae) were used for the identification, structural and functional characterization of FTs of the NAC family in soybean and cowpea transcriptomes. 50 and 42 NAC FTs were identified in the soybean and cowpea transcriptomes, respectively, submitted to root dehydration as compared to a non-stressed control. The proteins corresponding to the candidate genes were evaluated for the presence of conserved domains, being classified into five groups (NAC-*a* until NAC-*e*), with NAC-*d* being the most represented, while NAC-*c* presented only five sequences of cowpea and none of soybean. The proteins presented relative conservation, especially in the 5' region alongside a greater variation in the 3' region, reflecting on the secondary structure revealed by the three-dimensional modeling applied to nine proteins (five from cowpea and four from soybean). Analysis of NAC FTs expression revealed some induced gene candidates in all tissues and situations (including non-stressed controls), while others had low expression or specificity of time and tissue. The results indicate a significant diversity and distinct strategies regarding these factors in the analyzed legumes, being potentially useful for future use as molecular markers or in transgenesis, aiming at better performance under biotic and abiotic stresses.

Keywords: Abiotic stress. Cowpea. Soybean (*Glycine max*). Root dehydration. Bioinformatics.

LISTA DE FIGURAS

- Figura 1 — Representação esquemática dos motivos conservados das sequências proteicas de NAC do feijão-caupi (*V. unguiculata*) obtidos através da ferramenta MEME. Cada motivo é representado por uma cor e um número. As linhas pretas abaixo das caixas coloridas representam as regiões não conservadas das sequências..... 45
- Figura 2 — Representação esquemática dos motivos conservados das sequências proteicas de NAC da soja (*G. max*) obtidos através da ferramenta MEME. Cada motivo é representado por uma cor e um número. As linhas pretas abaixo das caixas coloridas representam as regiões não conservadas das sequências..... 46
- Figura 3 — As sequências proteicas (*full-length*) de soja (Glyma = 50 sequências) e feijão-caupi (TR = 42) foram alinhadas no Clustal Omega e a árvore fenética foi construída no programa MEGA6 através do método de Neighbor-Joining (NJ) com valor de bootstrap fixado em 1000. Cada subfamília NAC (a – e) está evidenciada por caixas coloridas: Vermelho – NAC-a; Verde – NAC-b; Amarelo – NAC-c; Azul – NAC-d; Lilás – NAC-e..... 56
- Figura 4 — Edição do alinhamento múltiplo de algumas das sequências proteicas de NAC de soja (Gm), feijão-caupi (Vu), Medicago (Mt), feijão comum (Pv), amendoim (Ah) e Arabidopsis (At). Os resíduos vermelhos são comuns a no mínimo metade das sequências. Os resíduos azuis são quimicamente similares em mais da metade das sequências. O alinhamento corresponde à região com maiores níveis de conservação/similaridade entre as sequências analisadas..... 57
- Figura 5 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-a demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME..... 58
- Figura 6 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-b demonstrando a similaridade da composição dos motivos

	conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME.....	58
Figura 7 —	Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-c demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de feijão-caupi (TR) com o auxílio da ferramenta MEME.....	59
Figura 8 —	Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-b demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME.....	59
Figura 9 —	Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-b demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME.....	59
Figura 10 —	Modelagem tridimensional das proteínas VuNAC-a, VuNAC-b, VuNAC-c, VuNAC-d e VuNAC-e gerado a partir do programa SWISS-MODEL com os seus respectivos gráficos Ramachandran.....	63
Figura 11 —	Modelagem tridimensional das proteínas GmNAC-a, GmNAC-b, GmNAC-d e GmNAC-e gerado a partir do programa SWISS-MODEL com seus respectivos gráficos Ramachandran.....	64
Figura 12 —	Perfil de expressão <i>in silico</i> (heatmap) para as sequências de soja e feijão-caupi obtidos após clusterização hierárquica realizada pelo programa Cluster 3.0.....	65

LISTA DE TABELAS

Tabela 1 —	Número de <i>seed sequences</i> proteicas dos fatores de transcrição da família NAC (NAC FTs) anotadas, com base na literatura, para cada organismo de referência depositadas no Genbank – NCBI e quantidade de sequências de transcritos obtidos via tBLASTn contra o banco de Fabaceae do GenBank – NCBI (<i>cut-off</i> ≤ e-05).....	41
Tabela 2 —	Relação dos motivos conservados mais abundantes identificados nas sequências de soja com o auxílio da ferramenta MEME.....	42
Tabela 3 —	Relação dos motivos conservados mais abundantes identificados nas sequências de feijão-caupi com o auxílio da ferramenta MEME.....	42
Tabela 4 —	Relação dos genes codificadores para NAC encontrados no genoma da soja (<i>Glycine max</i>).....	47
Tabela 5 —	Relação dos genes codificadores para NAC encontrados no genoma do feijão-caupi (<i>Vigna unguiculata</i>).....	50
Tabela 6 —	Primers desenhados para soja (Gm_NAC) e feijão-caupi (Vu_NAC) a partir das sequências obtidas nos bancos de dados de RNA-Seq de ambas as espécies , incluindo nome do primer, sequência de entrada, par do primer, posição de início, tamanho do Amplicon, temperatura de anelamento, conteúdo relativo de GC e as sequências.....	66

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

aa	Aminoácidos
ABA	Ácido Abscísico
Ah	<i>Arachis hypogaea</i>
At	<i>Arabidopsis thaliana</i>
ATAF	<i>Arabidopsis thaliana Activation Factor</i>
BACs	Cromossomos Artificiais de Bactéria (<i>Bacterial Artificial Chromosomes</i>)
BLAST	<i>Basic Local Alignment Search Tool</i>
CAM	Metabolismo Ácido das Crassuláceas (<i>Crassulacean Acid Metabolism</i>)
CO ₂	Gás Carbônico
CONAB	Companhia Nacional de Abastecimento
CUC	<i>Cup-shaped Cotyledon</i>
DNA	Ácido Desoxirribonucleico
DREB	<i>Dehydration Responsive Element Binding Protein</i>
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EST	<i>Expressed Sequence Tags</i>
FPKM	<i>Fragments per Kilobase Million</i>
FTs	Fatores de Transcrição
GenBank	Banco de dados genéticos
GENOSOJA	Consórcio Nacional da Soja
Gm	<i>Glycine max</i>
H ₂ O ₂	Peróxido de Hidrogênio
ha	Hectare
MEME	<i>Multiple Expectation Maximization for Motif Elicitation</i>
Mt	<i>Medicago truncatula</i>
NAC	<i>NAM, ATAF1/2, CUC2</i>
NAM	<i>No Apical Meristem</i>
NCBI	<i>National Center for Biotechnology Information</i>
NGS	<i>Next Generation Sequencing</i>
NordEST	Consórcio do Feijão-caupi

O ₂	Oxigênio
OH	Hidroxila
ORF	<i>Open Reading Frame</i>
pb	Pares de Base
PCR	Reação em Cadeia da Polimerase (<i>Polymerase Chain Reaction</i>)
PDB	<i>Protein Database</i>
Pfam	<i>Protein Families Database</i>
Pv	<i>Phaseolus vulgaris</i>
rDNA	DNA Ribossomal
ReNorBio	Rede Nordestina de Biotecnologia
RNA	Ácido Ribonucleico
RNA-Seq	Sequenciamento de RNA
ROS	<i>Reactive Oxygen Species</i> – Espécies Reativas de Oxigênio
RT-qPCR	PCR Quantitativa em Tempo Real (<i>Real-Time quantitative PCR</i>)
SAGE	Análise Serial da Expressão Gênica (<i>Serial Analysis of Gene Expression</i>)
Vu	<i>Vigna unguiculate</i>

SUMÁRIO

1	INTRODUÇÃO.....	17
1.1	OBJETIVOS.....	18
1.1.1	Objetivo geral.....	18
1.1.2	Objetivos específicos.....	19
2	REVISÃO BIBLIOGRÁFICA.....	20
2.1	A FAMÍLIA FABACEAE E SUA IMPORTÂNCIA PARA A AGRICULTURA...	20
2.2	A CULTURA DO FEIJÃO-CAUPI, DA SOJA E DE <i>STYLOSANTHES</i> NO BRASIL E NO MUNDO.....	20
2.3	O DÉFICIT HÍDRICO E SALINIDADE: IMPACTOS NA AGRICULTURA.....	22
2.4	MECANISMOS FISIOLÓGICOS, BIOQUÍMICOS E MOLECULARES DE ENFRENTAMENTO DE ESTRESSES EM PLANTAS.....	23
2.5	FATORES DE TRANSCRIÇÃO (FTS) EM PLANTAS.....	25
2.5.1	Principais famílias de Fatores de Transcrição na Defesa Vegetal.....	27
2.5.2	A Família NAC: estrutura, evolução e diversidade funcional.....	28
2.6	BIOINFORMÁTICA E PRINCIPAIS FERRAMENTAS DE ANÁLISE.....	30
2.7	BANCOS DE DADOS DE CAUPI (NORDESTE) E DA SOJA (GENOSOJA).....	32
3	ESTRUTURA E FUNÇÃO DE FATORES DE TRANSCRIÇÃO DA FAMÍLIA NAC EM SOJA E FEIJÃO-CAUPI SOB DESIDRATAÇÃO RADICULAR: UMA ANÁLISE COMPARATIVA.....	35
4	CONCLUSÕES.....	73
	REFERÊNCIAS.....	74

1 INTRODUÇÃO

A agricultura brasileira tem avançado significativamente no que diz respeito à produção e produtividade, mais especificamente nas culturas de importância agrícola, sendo um importante vetor para o desenvolvimento da economia brasileira. Contudo, tal avanço tem sido constantemente ameaçado tanto por conta de estresses bióticos (ex. vírus, nematoides e insetos), quanto por estresses abióticos, como seca, salinidade, temperaturas extremas, toxicidade química e estresse oxidativo, entre outros (GAO *et al.*, 2007; BUAINAIN e GARCIA, 2015). Dentre os estresses abióticos, as perdas no rendimento médio das colheitas em decorrência a seca podem passar dos 50% (GAO *et al.*, 2007) ou, até mesmo compreender até 70% (KAUR *et al.*, 2008; THAKUR *et al.*, 2010) da produtividade das culturas alimentares de base), representando um grande desafio diante de uma crescente população mundial.

Dentre as culturas agrícolas que sofrem grande impacto pela seca, encontram-se a soja (*Glycine max*) e o feijão-caupi (*Vigna unguiculata*), ambas pertencentes à família Fabaceae, a terceira maior família de angiospermas (LEWIS *et al.*, 2005). A família Fabaceae possui um papel importante, principalmente quanto a sua exploração econômica e para a recuperação de áreas degradadas (FRANCO e FARIA, 1997; SPRENT, 2001), visto que algumas espécies de leguminosas, através da simbiose com bactérias do gênero *Rhizobium*, são capazes de realizar a fixação biológica de nitrogênio do solo (GALLOWAY *et al.*, 2004; DELLA COLETTA, 2010).

Com o avanço tecnológico e o surgimento de novas plataformas de sequenciamento NGS (*Next Generation Sequencing*) muitos projetos de genômica funcional têm gerado um número expressivo de dados, especialmente no que se refere à identificação de genes importantes que atuam nas vias de defesa vegetal, transdução de sinais e regulação da expressão gênica durante estresses bióticos e abióticos (BARBA *et al.*, 2014; UNAMBA *et al.*, 2015). Neste sentido, a busca por plantas mais tolerantes aos diferentes tipos de estresse, principalmente através da transgenia, apresenta-se como uma alternativa promissora para a introdução de genes de interesse de outras espécies em culturas de interesse comercial (YAMAGUSHI-SHINOZAKI, 2000; RONDE, 2004; MITTLER, 2006).

Dentre os principais genes participantes das vias de resposta a estresses, destacam-se aqueles que codificam enzimas envolvidas na biossíntese de vários osmoprotetores, enzimas detoxificantes, genes envolvidos na captura e transporte de água e íons ou, ainda, genes que codificam os Fatores de Transcrição – FTs - proteínas envolvidas nas etapas iniciais de

modulação da expressão e transdução de sinais (BENKO-ISEPPON *et al.*, 2011; AMORIM *et al.*, 2017).

FTs coordenam diversos aspectos do desenvolvimento e crescimento vegetal, bem como aqueles relacionados às respostas às mudanças ambientais. Nesse contexto, a utilização de FTs tem se mostrado uma das alternativas promissoras na obtenção de plantas mais tolerantes à seca (SOUER *et al.*, 1996; AINDA *et al.*, 1997; YAMAGUSHI-SHINOZAKI, 2002; BECK *et al.*, 2007).

Dentre os diversos tipos de FTs conhecidos, encontram-se os membros da família NAC, que regulam inúmeras funções biológicas, sendo exclusivos dos vegetais (AIDA *et al.*, 1997; OLSEN *et al.*, 2005). A família NAC compreende uma variedade de proteínas vegetais compostas por um domínio NAC, N-terminal altamente conservado, acompanhado por diversos domínios C-terminais. O nome NAC deriva dos três primeiros genes descritos: *NAM* (*No Apical Meristem*), *ATAF1/2* (*Arabidopsis Transcription Activation Factor 1/2*) e *CUC2* (*cup-shaped cotyledon*) (SOUER *et al.*, 1996; AIDA *et al.*, 1997).

As proteínas NAC são bastante difundidas em plantas. Por exemplo, só no genoma de *Arabidopsis thaliana* estima-se que existam pelo menos cem genes codificantes para NAC (THE ARABIDOPSIS GENOME INITIATIVE, 2000; RIECHMANN *et al.*, 2000). Além de desempenhar controle transcricional em diversos processos metabólicos e fisiológicos das plantas, estas proteínas também atuam na resposta a estresses e infecções virais (XIE *et al.*, 1999; XIE *et al.*, 2000; REN *et al.*, 2000; COLLINGE e BOLLER, 2001). Diversas destas proteínas foram descritas como ativadores transcricionais (DUVAL *et al.*, 2002; ERNST *et al.*, 2004).

O presente estudo visou à identificação de NAC FTs em soja e feijão-caupi, inferindo sobre seu papel, abundância, diversidade, estrutura, além de sua expressão diferencial sob desidratação radicular, colaborando para a identificação de candidatos com potencial biotecnológico e no melhoramento destas leguminosas.

1.1 OBJETIVOS

1.1.1 Objetivo geral

- Avaliar a diversidade, estrutura e expressão de fatores de transcrição NAC (NAC FTs) envolvidos nos mecanismos moleculares em feijão-caupi e soja quando submetidas a estresse por desidratação radicular, inferindo sobre seu papel no estresse em questão,

bem como em sua diversidade e evolução, comparativamente às demais Angiospermas analisadas.

1.1.2 Objetivos específicos

- Identificar, quantificar e caracterizar os NAC FTs presentes no feijão-caupi (NordEST) e na soja (GENOSOJA) comparativamente entre ambas as culturas e a outras leguminosas já amostradas;
- Localizar os domínios e motivos conservados para estudos comparativos com sequências de genes previamente caracterizados, inferindo sobre suas características funcionais;
- Estabelecer um perfil de expressão *in silico* dos genes candidatos (a partir dos dados de RNA-Seq), visando à identificação de genes-candidatos associados à tolerância ou susceptibilidade ao estresse abiótico em tela;
- Gerar um banco local informatizado com os NAC FTs envolvidos nas respostas a estresses abióticos, disponibilizando-o para uso em projetos de melhoramento (seleção assistida e transgenia) das leguminosas analisadas.

2 REVISÃO BIBLIOGRÁFICA

2.1 A FAMÍLIA FABACEAE E SUA IMPORTÂNCIA PARA A AGRICULTURA

A família das leguminosas, ou Fabaceae, representa a terceira maior família de angiospermas existente no mundo, com mais de 19.327 espécies e 727 gêneros compreendidos em 36 tribos (LEWIS *et al.*, 2005). Ocorre em praticamente todas as formações vegetais do planeta, com maior concentração e número de endemismos nas regiões neotropicais (LAVIN *et al.*, 2004). Pertence à ordem Fabales (APG III, 2009), e está dividida em três subfamílias: Caesalpinioideae, Mimosoideae e Papilionoideae (POLHILL *et al.*, 1981). Dentre essas, a subfamília Caesalpinioideae se destaca como um grupo parafilético do qual se originaram as outras duas subfamílias (LPWG, 2013).

Na flora brasileira, a família Fabaceae é a mais diversa, inclusive na Caatinga, principal região natural do nordeste brasileiro (AMORIM *et al.*, 2016). Em toda a extensão do território nacional, com 2.807 espécies agrupadas em 22 gêneros, dos quais 15 são endêmicos, compreendendo a família mais bem representada, sendo encontrada em quase todos os biomas e ecossistemas do país (BFG 2015 – *The Brazil Flora Group*).

A subfamília Papilionoideae (ou Faboideae), é constituída por 430 gêneros e, aproximadamente, 12.600 espécies de ampla distribuição pelo mundo. É considerada a subfamília mais derivada dentre as leguminosas, e também a de maior importância econômica, incluindo a soja (*Glycine max*), a ervilha (*Pisum sativum*), o feijão comum (*Phaseolus vulgaris*), o feijão-caupi (*Vigna unguiculata*), a alfafa (*Medicago sativa*), o grão-de-bico (*Cicer arietinum*), o amendoim (*Arachis hypogaea*), entre outros. A subfamília ainda é importante para a indústria madeireira, química, farmacêutica, etc. (LEWIS *et al.*, 2005).

Alguns membros da família Fabaceae, incluindo o feijão-caupi (*Vigna unguiculata*), são de grande importância agronômica, visto que são capazes de realizar, em associação com bactérias do gênero *Rhizobium*, promovendo a fixação de nitrogênio do solo através da formação de nódulos radiculares (GALLOWAY *et al.*, 2004; DELLA COLETTA, 2010).

2.2 A CULTURA DO FEIJÃO-CAUPI, DA SOJA E DE *STYLOSANTHES* NO BRASIL E NO MUNDO

O feijão-caupi [*Vigna unguiculata* (L.) Walp], também conhecido como feijão-de-corda ou feijão-macassar é uma das culturas mais tradicionais da região nordeste do Brasil, de

importância para a agricultura familiar. Por ser uma excelente fonte de proteínas (25 % em média) e apresentar todos os aminoácidos essenciais, carboidratos (62% em média), vitaminas e minerais, não conter colesterol e possuir baixa quantidade de gordura (ANDRADE JÚNIOR *et al.*, 2003), esta cultura figura como uma das principais fontes de alimento da população de baixa renda do nordeste brasileiro, além de também ser utilizada na alimentação animal, como forrageira para adubo verde e cobertura do solo (CORDEIRO, 1998; ANDRADE JÚNIOR *et al.*, 2003; EMBRAPA, 2011).

Dentre as leguminosas cultivadas, o feijão-caupi figura em posição de destaque em nível mundial graças à sua rusticidade, com uma área de cultivo estimada em 12,5 milhões de hectares (ha) e produção anual superior a 5 milhões de toneladas. Mais da metade da área total de cultivo do feijão-caupi (cerca de 60 %, ou 8 milhões de hectares) está situada nas áreas central e oeste do continente Africano. O restante encontra-se distribuído pela América do Sul, América Central e Ásia, com alguns pequenos focos no sudoeste Europeu, sudoeste dos Estados Unidos e Oceania (FREIRE FILHO, 2011).

No Brasil, a produção de feijão-caupi concentra-se, principalmente, na região Nordeste e Norte e está se expandindo para a região Centro-Oeste, principalmente para o Estado de Mato Grosso (FREIRE-FILHO, 2011), em virtude de sua adaptabilidade elevada às condições tropicais, ao custo relativamente baixo de produção e ao fácil manejo (EMBRAPA, 2011). Segundo dados divulgados pela CONAB (2016), em todo o território nacional, a área total cultivada de feijão foi de aproximadamente 2.934 mil ha com uma produção de aproximadamente 3 milhões de toneladas.

Outra leguminosa de maior destaque no cenário agrícola é a soja (*Glycine max* (L.) Merrill), sendo uma das culturas mais importantes na economia mundial. Dentre as diversas utilizações de seus grãos pela agroindústria, destacam-se a produção de óleo vegetal e rações para alimentação animal, sendo também muito utilizada na indústria química e de alimentos, bem como pela indústria de biocombustíveis (COSTA NETO e ROSSI, 2000). Segundo o levantamento da safra brasileira de grãos 2015/2016 (CONAB, 2016), a cultura da soja é responsável por 57 % da área total cultivada no território brasileiro, ocupando aproximadamente 58,17 milhões de ha e responsável por uma produção de cerca de 96,5 milhões de toneladas na safra 2015/2016.

Dentre as Fabaceae tropicais destaca-se o gênero *Stylosanthes*, um grupo megatérmico e pantropical com cerca de 50 espécies descritas, amplamente distribuídas no continente americano e, como resultado do processo evolutivo e ecotipos submetidos às diferentes condições de clima, solos e estresses bióticos, destacando-se por apresentar grande variação

morfológica (LEWIS *et al.*, 2002; KARIA *et al.*, 2002; SILVESTRE, 2013). Seu porte é prostrado a ereto, podendo alcançar a altura de 1,5 metros. Apresenta folhas trifolioladas, flores pequenas e caracteriza-se pela grande diversidade agronômica (FERREIRA e COSTA, 1979; SCHULTZE-KRAFT *et al.*, 1984; STACE e EDYE, 1984; SILVESTRE, 2013).

As espécies de *Stylosanthes* apresentam como centro de origem as Américas Central e do Sul, contudo, em decorrência de seu uso para forragem, são encontradas grandes áreas de cultivo também nas regiões subtropicais das Américas e principalmente na Austrália. A maioria das espécies é perene, com potente sistema radicular e alta tolerância à seca. Devido à sua grande capacidade de adaptação a solos de baixa fertilidade e simbiose com bactérias fixadoras de nitrogênio, as espécies de *Stylosanthes* apresentam grande capacidade colonizadora (STACE e EDYE, 1984; ANDRADE e KARIA, 2000; SILVESTRE, 2013). Dentre as espécies de *Stylosanthes*, destacam-se *S. guianensis*, *S. macrocephala*, *S. capitata* e *S. scabra*, que são amplamente cultivadas em todo o mundo principalmente para a produção de matéria seca para a alimentação animal. O “Estilosantes Campo Grande”, por exemplo, é comercializado como uma mistura de sementes de *S. capitata* e *S. macrocephala*, na proporção de 80% e 20% respectivamente (EMBRAPA, 2007).

2.3 O DÉFICIT HÍDRICO E SALINIDADE E SEUS IMPACTOS NA AGRICULTURA

Dentre os principais tipos de estresse abiótico que mais afetam mundialmente a produção das culturas de interesse agrícola, destacam-se a seca, a salinidade e temperaturas extremas (MANTRI *et al.*, 2012). Tais fatores de estresse podem causar perdas entre 50 -70 % do rendimento de culturas importantes, afetando, assim, o abastecimento de alguns tipos de alimento em todo o mundo (VORASOOT *et al.*, 2003; KAUR *et al.*, 2008; THAKUR *et al.*, 2010).

Mudanças climáticas ocasionadas pelo aumento massivo nos níveis de emissão de “gases do efeito estufa”, como o aquecimento global, têm exposto determinadas regiões a condições cada vez mais secas e quentes, fator esse que tem contribuído para uma retração na extensão das áreas destinadas à agricultura (HURA *et al.*, 2007; STILLER *et al.*, 2008).

O déficit hídrico pode ser definido como todo o conteúdo de água de um tecido ou célula que está abaixo do conteúdo de água mais alto exibido no maior estado de hidratação, quadro comumente constatado em plantas submetidas à seca, onde a disponibilidade hídrica do solo diminui (VASELLATI, 2001; TAIZ e ZEIGER, 2006).

De acordo com Bray (2004), as perdas ocasionadas em decorrência da seca podem comprometer em mais de 50% a produção mundial das culturas de interesse agrícola, enquanto os prejuízos ocasionados por estresses bióticos, como nematoides e viroses, reduzem entre 10% e 20% da produtividade destas. Durante a fase de germinação, por exemplo, o déficit hídrico pode reduzir em 20% a produtividade da soja, com danos ainda maiores no florescimento, chegando a 64% de comprometimento desta cultura.

Na cultura da soja, a disponibilidade hídrica é de vital importância, principalmente em dois períodos do desenvolvimento: a germinação-emergência e a floração-enchimento de grãos. No primeiro estágio, quando a disponibilidade hídrica é abaixo da ideal, prejudica a obtenção de uma uniformidade na população de plantas. No segundo estágio, quando sob níveis críticos, podem ocorrer alterações fisiológicas, como o fechamento estomático e o enrolamento das folhas. Em decorrência dessas respostas ao déficit hídrico, ocorre a queda prematura de folhas e flores e o abortamento de vagens, resultando na redução da produtividade e da cultura. Durante a fase vegetativa, a deficiência hídrica provoca a redução na taxa de crescimento e área foliar, enquanto que na fase reprodutiva, além das reduções mais drásticas no rendimento, um adiantamento na maturação, o que resulta em um menor acúmulo de matéria seca, vagens vazias e chocamento de grãos (MAHMOOD, 199; SAITOH, 1999; BONATO, 2000; SALINET, 2009).

Da mesma forma que para as demais culturas de importância agrícola, o feijão-caupi também tem sido afetado pela seca. Estima-se que, anualmente, a cultura deste grão sofre perdas de mais de 300 mil toneladas devido à seca, somente na África, continente que abriga os principais países produtores mundiais (RAMALHO *et al.*, 2009). Embora o caupi seja considerado moderadamente tolerante à seca, da mesma forma que para soja, o déficit hídrico é especialmente prejudicial nos três principais estádios do desenvolvimento, como a germinação, a floração e o enchimento de grãos (SORATTO *et al.*, 2003). Apesar disso, cerca de 60% do cultivo de feijão-caupi no mundo encontra-se em áreas de deficiência hídrica constante, cujo fator tem sido responsável pela baixa produtividade (AGUIAR *et al.*, 2008; BEEBE *et al.*, 2008; VALE *et al.*, 2012).

2.4 MECANISMOS FISIOLÓGICOS, BIOQUÍMICOS E MOLECULARES DE ENFRENTAMENTO DE ESTRESSES EM PLANTAS

A série de mudanças causadas pelos estresses bióticos e abióticos nas plantas, como a seca, é de ordem fisiológica, bioquímica, morfológica e molecular (WANG *et al.*, 2001; LEI *et*

al., 2006), afetando seu metabolismo, crescimento e desenvolvimento (BARTELS e SUNKARS, 2005). Como resposta ao déficit hídrico, as plantas podem desenvolver mecanismos tanto para retardar ou tolerar a desidratação, pela manutenção da hidratação do tecido, e pela capacidade de sobreviver mantendo suas funções enquanto desidratada (TAIZ e ZEIGER *et al.*, 2006).

Alguns dos mecanismos fisiológicos, desenvolvidos pelas plantas ao longo do processo evolutivo, com a finalidade de retardar a desidratação são, por exemplo, além do fechamento estomático, o desenvolvimento de tricomas, a redução da área foliar e perda das folhas velhas e/ou danificadas, enrolamento das folhas, dentre outros. Por outro lado, as plantas podem procurar maximizar a absorção de água através do crescimento radicular mais profundo (ISODA e WANG, 2002; TAIZ e ZEIGER, 2006; BARNABÁS *et al.*, 2008). Além disso, algumas plantas, como o caso da *Mesembryanthemum crystallinum*, possuem a habilidade de mudar o seu regime fotossintético de C₃ para o CAM (Metabolismo Ácido das Crassuláceas) em resposta ao estresse salino ou a seca (SAIRAN e TYAGI, 2004).

Grande parte dos mecanismos bioquímicos e moleculares é desencadeada já nos primeiros momentos de percepção do estresse (SANTOS e CARLESSO, 1998), muitos dos quais envolvem o ajustamento osmótico (acúmulo de solutos pela célula) e eficiência na remoção de espécies reativas de oxigênio (ROS, *Reactive Oxygen Species*) geradas em consequência dos danos causados pelo estresse (TAIZ e ZEIGER, 2006). O ajuste osmótico é o processo pelo qual se reduz o potencial hídrico da célula, a fim de se manter o equilíbrio hídrico no interior celular, ou seja, seu volume e turgor (TAIZ e ZEIGER, 2006). Quando submetidas à seca, a planta acumula vários metabólitos que passam a atuar como osmólitos, não apenas na manutenção do turgor da célula, mas também na detoxificação de ROS e estabilização de proteínas e estruturas celulares (RUIZ-LOZANO, 2003; CHINNUSAMY *et al.*, 2005). Entre estes osmólitos estão açúcares (oligossacarídeos da família da rafinose, sacarose, trehalose e sorbitol), manitol, prolina, algumas proteínas e aminas (TAJI, *et al.*, 2002; BARTELS e SUNKAR, 2005; SEKI *et al.*, 2007).

O ajustamento osmótico tem sido considerado um dos processos cruciais na adaptação das plantas à seca, por sustentar a atividade metabólica no tecido permitindo a retomada do crescimento após a reidratação, embora tal capacidade varie entre espécies e genótipos (SERRAJ e SINCLAIR, 2002; CHAVES *et al.*, 2003; SALINET, 2009).

Durante a seca, por exemplo, a disponibilidade de água no solo diminui e o *continuum* solo-planta-atmosfera pode ser perdido (VASELLATI *et al.*, 2001). Nesse contexto, a planta fecha seus estômatos com o objetivo de minimizar a perda de água para o ambiente (OHASHI

et al., 2006). O fechamento estomático, sob tais condições, se dá pelo aumento do ácido abscísico - ABA e como consequência, limita a difusão de CO₂ de espaços intercelulares para o interior do cloroplasto, diminuindo assim a fotossíntese (CORNIC, 2000; LIMA *et al.*, 2002). Essa percepção se dá pelas raízes, ao perceberem a diminuição da disponibilidade de água no solo, promovendo um acúmulo de ABA nos estômatos. Posteriormente, o ABA é redistribuído para os demais tecidos da planta (via xilema), agindo como um sinalizador químico para o fechamento dos estômatos e induzindo a expressão de genes relacionados ao estresse (LIU *et al.*, 2005; SEKI *et al.*, 2007).

Outro fator que induz o fechamento estomático sob regime de seca é a redução do potencial hídrico das folhas e da turgescência, que culmina nesse evento. Alguns autores demonstraram em seus trabalhos evidências que combinam ambos os tipos de sinalização, química e fisiológica, atuando juntas ou em tempos distintos no fechamento estomático (ASAMAA *et al.*, 2002; COMSTOCK, 2002; KALEFETOGLU e EKMEKÇI, 2005).

Quando ocorre a limitação da fotossíntese, ocorre também o consumo de elétrons liberados da água, ocasionando o excesso de energia de excitação, a qual, se não for eliminada de forma eficiente, pode ocasionar danos (MULLER *et al.*, 2001; LAWLOR, 2002). Quando o estado excitado da clorofila não é rapidamente dissipado pela transferência de excitação ou fotoquímica, ou seja, quando a energia armazenada não é dissipada, ela pode reagir com o oxigênio formando as ROS, que incluem o singlete (¹O₂), ânions radicais de superóxidos (O₂⁻), radicais hidroxilas (OH) e peróxidos de hidrogênio (H₂O₂) (BARTLES, 2001; CHAVES *et al.*, 2003; APEL e HIRT, 2004).

2.5 FATORES DE TRANSCRIÇÃO (FTS) EM PLANTAS

Por serem organismos sésseis, as plantas encontram-se constantemente submetidas a condições ambientais oscilantes e diversos fatores de estresses bióticos, tais como vírus, bactérias, herbivoria, e abióticos, como déficit hídrico, salinidade, temperaturas extremas, e radiação. Estes inúmeros fatores podem ocorrer em diversos estágios de desenvolvimento da planta e, normalmente, em condições naturais, os vegetais são acometidos por mais de um tipo de estresse simultaneamente, onde essa combinação de estresses atua como fator limitante restringindo o crescimento e o desenvolvimento da planta (CHEN *et al.*, 2012).

Em plantas, o controle transcricional da expressão de genes responsivos a estresse figura um papel crucial no que diz respeito ao arcabouço molecular de defesa contra estresses bióticos e abióticos. Diversos estudos realizados ao longo dos anos têm identificado e caracterizado

inúmeros fatores essenciais na modulação e regulação dos mecanismos de resposta das plantas a estes estresses (SINGH *et al.*, 2002).

Ao longo da evolução, os organismos vegetais desenvolveram mecanismos e estratégias de enfrentamento e resistência a estresses bióticos e abióticos, cuja eficiência depende de uma combinação de diversos processos fisiológicos, bioquímicos e moleculares que vão atuar desde a percepção do estresse e consequente desencadeamento das vias de sinalização e transdução de sinais, bem como de moléculas reguladoras (URANO *et al.*, 2010). Por exemplo, a resistência da planta frente ao patógeno é acompanhada pelo desencadeamento (rápido) de uma resposta de defesa composta de múltiplos componentes, como a resposta hipersensível (HR), o arsenal químico composto, por exemplo, por fitoalexinas e enzimas hidrolíticas, modificações estruturais, tais como paredes celulares ricas em lignina e hidroxiprolina (DIXON e HARRIS, 1994). Estes mecanismos são regulados por múltiplas vias de sinalização. Diversos estudos demonstram a sobreposição entre os padrões de expressão de diversos genes que compõem essas vias e sendo induzidos em resposta a diversos tipos de estresses (GLAZEBOOK, 2001; KNIGHT e KNIGHT, 2001; SEKI *et al.*, 2001; CHEN *et al.*, 2002).

A indução de genes relacionados às vias de resposta e enfrentamento a estresses ocorrem, primariamente, em nível transcricional, desempenhando papel essencial na regulação tempo-espacial do perfil de expressão de genes estresse-específicos (RUSHTON e SOMSSICH, 1998, SINGH *et al.*, 2002). Alguns destes genes são fundamentais na defesa vegetal, a exemplo daqueles que codificam enzimas envolvidas na biossíntese de osmoprotetores, enzimas detoxificantes, genes envolvidos no transporte de água e íons, ou aqueles que codificam proteínas envolvidas nas etapas iniciais de expressão e regulação e transdução de sinais em resposta aos estresses, como os Fatores de Transcrição (SEKI *et al.*, 2002; BENKO-ISEPPON *et al.*, 2011).

Os FTs são componentes fundamentais do sistema de imunidade inata e defesa basal de plantas, participando na regulação da expressão gênica através da ligação de sequências específicas de DNA na região regulatória dos genes alvos, modulando a atividade da RNA-polimerase no processo transcricional (SOUER *et al.*, 1996; AIDA *et al.*, 1997; ZHENG *et al.*, 2016). Pertencem a grandes famílias gênicas, algumas das quais específicas de plantas, a exemplo a família NAC (RIECHMANN *et al.*, 2000).

A relação entre os FTs com aquisição de tolerância a diversos tipos de estresses em plantas tem despertado grande interesse em estudos genéticos, principalmente pelo seu potencial uso biotecnológico no melhoramento de plantas. (PELLEGRINESHI *et al.*, 2004; KASUGA, 2004; SALINET, 2009).

2.5.1 Principais famílias de Fatores de Transcrição na Defesa Vegetal

A família de fatores responsivos ao etileno – ERF (*Ethylene Responsive Factors*) pertencem à superfamília AP2/ERF, um dos maiores grupos de FTs de plantas, caracterizados pela presença de um domínio de ligação ao DNA AP2 (RIECHMANN *et al.*, 2000). Geralmente, a família ERF é classificada em duas grandes subfamílias, DREB e ERF, com base no domínio de ligação ao DNA (SAKUMA *et al.*, 2002).

A subfamília DREB (*Dehydration Responsive Element Binding protein*), atua no topo da cascata de eventos moleculares que culminam em resposta à desidratação celular (YAMAGUSHI-SHINOZAKI, 2002). Estudos com *A. thaliana* demonstraram que os fatores DREB atuam na indução de genes que participam das vias de resposta à seca (PELLEGRINESHI *et al.*, 2004), embora diversas outras famílias também tenham sido propostas desempenhar a mesma função (SALINET, 2009). Além de mediar mecanismos de resposta cruzada (*crosstalk*) entre os genes de defesa vegetal, membros da subfamília ERF também atuam nas vias de defesa a estresses bióticos, a exemplo do AtERF1 (ERF1 de *A. thaliana*), que confere resistência a uma série de fitopatógenos conhecidos (BERROCA-LOBO *et al.*, 2004).

FTs da família WRKY atuam nos processos de reconhecimento dos elementos W-box, com a sequência TTGACY (onde Y = C ou T), nas regiões promotoras dos patógenos ou genes responsivos aos padrões moleculares associados a patógenos - PAMPs (PAMPs, *Pathogen-Associated Molecular Pattern*, modulando positivamente as vias de resposta precoce contra diversos fitopatógenos (PANDEY *et al.*, 2009; RUSHTON *et al.*, 2010; ISHIHAMA *et al.*, 2012; AMORIM *et al.*, 2017). Esta família é caracterizada pela presença do domínio altamente conservado, WRKY-DBD, de 60 aminoácidos, que contém o motivo praticamente invariável WRKYGQK na região N-terminal (EULGEN *et al.*, 2000). A família WRKY é subdividida em três subfamílias (I, II e III) com base em relações filogenéticas, número de domínios WRKY e tipo do motivo dedo de zinco (*zinc finger*). A subfamília I apresenta dois domínios WRKY, enquanto as subfamílias II e III apresentam apenas um.

Outra importante família de FTs de plantas encontra-se a bZIP. Membros desta família são caracterizados pelo domínio conservado de, aproximadamente, 70 aminoácidos, que dá nome a família. Apresentam uma região básica (BR) e um zíper de leucina (LZ) funcionalmente distinto (HURST *et al.*, 1994). Os bZIP FTs são considerados reguladores-chave nas vias de sinalização mediadas por Ácido Salicílico, desempenhando papel crucial na defesa contra o ataque de patógenos (DURRANT e DONG, 2004; PIETERSE *et al.*, 2012).

2.5.2 A Família NAC: estrutura, evolução e diversidade funcional

Os FTs da família NAC (NAC FTs) são exclusivos de plantas, multifuncionais, atuando em diversos aspectos do desenvolvimento e crescimento vegetal, desde o desenvolvimento embrionário até a senescência, bem como nas vias de resposta a estresses bióticos e abióticos (AIDA *et al.*, 1997; OLSEN *et al.*, 2005; JENSEN e SKRIVER, 2014).

O domínio NAC foi identificado com base nas sequências consenso das proteínas NAM (*No Apical Meristem*) de *Petunia* e ATAF1/2 (*Arabidopsis Transcription Activation Fator 1/2*) e CUC2 (*Cup-shaped Cotyledon*) de *Arabidopsis*. As proteínas desta família apresentam um domínio NAC N-terminal altamente conservado, além de diversos domínios C-terminais (SOUER *et al.*, 1996; AIDA *et al.*, 1997).

Os membros da grande família NAC, associados ao estresse, mais bem caracterizados pertencem ao subgrupo III-3 (JENSEN *et al.*, 2010), também conhecido como NAC estresse-responsivos (SNAC) (NURUZZAMAN *et al.*, 2010), dos quais, três membros intimamente relacionados deste subgrupo, ANA019, ANAC055 e ANAC072 (RD26) são induzidos por seca, salinidade e os hormônios ácido abscísico (ABA), que é conhecido por sua participação no controle de diversos processos fisiológicos em plantas e pelo seu papel nas vias de resposta ao estresse abiótico, e ácido jasmônico (JA), que por sua vez é associado às vias de resposta ao estresse biótico (FUJITA *et al.*, 2006; JENSEN e SKRIVER, 2014).

Acredita-se que proteínas NAC se desenvolveram por volta de 400 milhões de anos atrás (ZHU *et al.*, 2012), sendo a maioria das subfamílias surgido antes da divergência entre monocotiledôneas e dicotiledôneas, com algumas subfamílias restritas a traqueófitas, monocotiledôneas, dicotiledôneas ou, raramente, específicas de algumas famílias (RUSHTON *et al.*, 2008; SHEN *et al.*, 2009; ZHU *et al.*, 2012; HUSSEY *et al.*, 2015).

NAC FTs são caracterizados pela presença do domínio NAC, que contém uma dimerização N-terminal e domínio de ligação ao DNA (DBD), “o domínio NAC propriamente dito”, e domínios regulatórios da transcrição (TRDs) variáveis na extremidade C-terminal, modulando a variação na regulação da atividade transcricional (YAMAGUCHI *et al.*, 2008; JENSEN *et al.*, 2010). Essa região altamente variável geralmente opera como domínio funcional, ativando ou reprimindo a transcrição (TRAN *et al.*, 2004; HU *et al.*, 2006; KIM *et al.*, 2007). Já, a região N-terminal contém um domínio altamente conservado de, aproximadamente, 150-160 aminoácidos, sendo dividida em cinco subdomínios (A – E) (Ooka *et al.*, 2003). A função do domínio NAC foi associado com a localização celular, ligação ao

DNA e a formação de homodímeros e heterodímeros com outras proteínas NAC contendo o mesmo domínio (OLSEN *et al.*, 2005).

Estudos recentes analisaram as interações entre o domínio NAC com um palíndromo do sítio de ligação de 7 pares de base de ANAC019 tanto através de técnicas de cristalografia, quanto em solução (OLSEN *et al.*, 2005; WELNER *et al.*, 2012). Ernst *et al.* (2004) identificaram uma nova dobra dimérica consistindo basicamente de folhas- β em análises estruturais por cristalografia dos 168 resíduos de aminoácidos do domínio NAC do FT ANAC019 de *A. thaliana* (ERNST *et al.*, 2004; JENSEN e SKRIVER, 2014). O domínio NAC insere uma folha- β externa com a sequência de aminoácidos WKATQTD altamente conservada na maior alça do DNA de forma específica ao mecanismo do fator de transcrição específico de plantas da família WRKY. Essa similaridade pode ser explicada tanto através da origem evolutiva comum, quanto através da convergência evolutiva destes fatores de transcrição. Vários resíduos de aminoácidos básicos são conservados e encontram-se dentro da distância de ligação ao DNA. Contudo, quando modificada uma Arg-88 por uma Alanina ocorre diminuição na afinidade de ligação do FT ao DNA, indicando que este resíduo (Arg-88) é essencial para a ligação NAC/DNA e, conseqüentemente, sua funcionalidade (WELNER *et al.*, 2012).

Grande parte dos membros da família NAC pode ser identificada com base em seus domínios conservados, os quais são subdivididos em subdomínios, cujo número pode variar de espécie para espécie. Cada subdomínio pode ser identificado e classificado de acordo com a composição específica de aminoácidos de seus respectivos motivos. Kikuchi *et al.* (2000) identificaram 5 subdomínios NAC em arroz (A – E), ricos em aminoácidos básicos (R, K e H). O que difere um subdomínio dos demais é a distribuição de aminoácidos positivos e negativos e suas relativas abundâncias nos motivos.

De forma geral, os NAC FTs envolvidos tanto na resposta a estresses quanto na senescência, ligam-se a uma região promotora com o motivo CGT[AG]. As diferentes sequências flangeadoras também se encontram em contato com o domínio NAC e estão intimamente ligadas com a afinidade e especificidade de ligação com o sítio de ligação ao promotor do alvo (WELNER *et al.*, 2012; JENSEN e SKRIVER, 2014). Contudo, há uma exceção em relação a *ATAF2*, que também é estresse induzido, pois este se liga a uma sequência não relacionada do promotor *NIT2* codificando um gene de síntese do hormônio auxina (DELESSERT *et al.*, 2005; HUH *et al.*, 2012).

2.6 BIOINFORMÁTICA E PRINCIPAIS FERRAMENTAS DE ANÁLISE

No início dos anos 90, devido ao enorme aumento na quantidade de sequências geradas a partir dos primeiros projetos de sequenciamento em larga escala, em especial, com o início do projeto Genoma Humano e subsequente disponibilização de plataformas de sequenciamento automáticos de DNA capazes de gerar dados genômicos em escalas nunca antes vistas, a informática angariou um lugar de destaque no campo da biologia com o objetivo de adaptar e otimizar os bancos de dados e ferramentas de análises a esta nova realidade, onde o volume de informações era cada vez maior. Desde então, milhares de sequências de nucleotídeos e aminoácidos são depositadas em bancos de dados como o GenBank (SANTOS e ORTEGA, 2002). Nascia então a bioinformática: um campo interdisciplinar que engloba diversas áreas de conhecimento, principalmente, a biologia e a informática para a geração, gerenciamento e processamento das informações. Esta ciência auxilia a administração e a interpretação dos dados oriundos de sequenciamento, fornecendo informações valiosas para estudos em diversas áreas, tais como genômica, transcriptômica e proteômica através de análises com programas e ferramentas computacionais de alto desempenho (CASSILHAS *et al.*, 2003; QUEIROZ, 2005; TEIXEIRA, 2005; PROSDOCIMI *et al.*, 2012).

Segundo de Araújo *et al.* (2008), a bioinformática foi, inicialmente, considerada um paradoxo entre o “mundo real e o imaginário”, uma vez que era inconcebível o emprego de técnicas e abordagens computacionais para a elucidação de questões biológicas. Contudo, ao longo dos últimos tornou-se uma área essencial nos estudos biológicos, principalmente devido aos avanços nas tecnologias de sequenciamento, necessitando de uma maior integração entre profissionais da biologia e da informática, preenchendo, assim, as lacunas que existiam e atendendo a uma demanda cada vez maior, com a geração de programas/ ferramentas cada vez mais específicas e adequadas para cada tipo de análise.

Para Cho e Walbot (2001), a descoberta de novos genes, bem como a determinação de suas funções, geralmente são baseadas na união de técnicas genéticas, bioquímicas e de biologia molecular, contudo, com o surgimento da bioinformática, este processo tem se tornado cada vez mais automatizado, otimizando a geração de processos e produtos, demandando menos recursos financeiros e fornecendo resultados em um intervalo consideravelmente reduzido. Nesse contexto, a bioinformática tem contribuído significativamente para avanços tanto na pesquisa básica, como na ciência aplicada, proporcionando a elucidação das bases moleculares envolvidas em importantes eventos celulares e assegurando sua aplicação direta em diversos

setores de interesse, como agricultura, indústria farmacêutica e ciências da saúde (DE ARAÚJO *et al.*, 2008).

Dentre as diversas ferramentas rotineiramente utilizadas na análise de dados biológicos é o BLAST (*Basic Local Alignment Search Tool*), pela qual são realizados alinhamentos locais de sequências de nucleotídeos ou aminoácidos, comparando-as com base em homologia. O BLAST utiliza uma aproximação direta em seus alinhamentos no que diz respeito às medidas de similaridade local, otimizando-as. Esta aproximação é denominada de MSP (*maximal segment pair*). Os resultados matemáticos obtidos referentes às propriedades estocásticas dos escores MSP permitem uma análise do desempenho deste método, bem como a significância estatística dos alinhamentos gerados a partir do BLAST. A utilização desta ferramenta e de seus algoritmos permite a construção de programas consideravelmente rápidos para a busca em bancos de dados. Adicionalmente, variações e implementações alternativas na ideia central do algoritmo utilizado na ferramenta BLAST permitem sua adaptação para diferentes contextos (ALTSCHUL *et al.*, 1990; ALTSCHUL *et al.*, 1997; YE *et al.*, 2006; JOHNSON *et al.*, 2008).

Outra etapa fundamental durante o processo de anotação e caracterização funcional é a identificação do domínio conservado (*Conserved Domain*). Domínios proteicos são unidades de evolução molecular distintas, normalmente associados à função molecular, como catálise ou ligação. Em geral, representam pequenas unidades da estrutura tridimensional da proteína. A identificação de domínios conservados funcionalmente caracterizados nas sequências proteicas promove os primeiros indícios em relação a sua possível função. Uma das ferramentas mais utilizadas para a identificação de domínios conservados é o CD-Search do NCBI, que utiliza o banco de dados de Domínios Conservados (*CDD – Conserved Domain Database*), o qual disponibiliza uma quantidade massiva de informações relacionadas aos domínios conservados, suas famílias, características estruturais e função celular e molecular (MARCHLER-BAUER *et al.*, 2002; MARCHLER-BAUER *et al.*, 2005).

Durante o processo de caracterização de sequências, a busca por motivos conservados é essencial, pois permite a identificação de pequenas sequências (de DNA ou aminoácidos) altamente conservadas dentro das famílias gênicas, o que permite uma melhor caracterização das sequências analisadas e de suas respectivas famílias gênicas. Com este objetivo, o MEME suite disponibiliza os principais programas necessários para tais buscas e análises (BAILEY *et al.*, 2009).

A estrutura tridimensional das proteínas é essencial para uma compreensão mais aprofundada acerca de suas bases moleculares bem como de suas funções. Combinando as informações das sequências com as estruturas 3D proporcionam dados valiosos para diversos

tipos de análises. Técnicas para a solução experimental de estruturas tridimensionais por cristalografia de Raios-X e espectroscopia por ressonância magnética nuclear avançaram consideravelmente nos últimos anos e foram amplamente utilizadas, sendo responsáveis pela elucidação da estrutura tridimensional de mais de 22.000 proteínas, cujos modelos e informações estão depositados no PDB (WESTBROOK *et al.*, 2003). Contudo, o uso destas técnicas para a resolução das estruturas tridimensionais das proteínas, além de serem consideravelmente onerosas, consomem um tempo relativamente elevado, além de não serem garantia de sucesso. Isso é refletido pelo fato de o número de proteínas estruturalmente caracterizadas ser aproximadamente mil vezes menor do que o número de sequências proteicas depositadas nos bancos de dados SwissProt e TrEMBL (BOECKMANN *et al.*, 2003). Como alternativa para contornar este fato, softwares de modelagem por homologia, como o SWISS-MODEL têm sido amplamente utilizados para a resolução da estrutura tridimensional de proteínas bem anotadas e caracterizadas cujos modelos encontram-se depositados nos principais bancos de dados proteicos existentes (EYRICH *et al.*, 2001; KOPP e SCHWEDE, 2003).

2.7 BANCOS DE DADOS DO FEIJÃO-CAUPI (NORDEST) E DA SOJA (GENOSOJA)

Uma das principais características das pesquisas genômicas atuais é a geração de uma quantidade massiva de dados brutos. À medida que essa quantidade de dados cresce, metodologias computacionais mais sofisticadas são necessárias para administrar o enorme fluxo de informações. Deste modo, um dos primeiros desafios da era genômica foi armazenar, gerenciar e lidar com um volume massivo de informações através do estabelecimento e uso de bancos de dados informatizados. O desenvolvimento de bancos de dados para a manipulação de vastas quantidades de dados biológicos é uma das principais atribuições da bioinformática (XIONG, 2006).

Em bioinformática, um banco de dados (*database*) é um repositório computadorizado de arquivos utilizado para armazenar e organizar as informações de forma que elas sejam recuperadas ou acessadas facilmente, através de um conjunto de critérios de busca. O objetivo principal do desenvolvimento de um banco de dados é organizar os dados de forma estruturada. Basicamente, cada registro (*entry*) deve conter características que o identifiquem as informações contidas neles, como por exemplo, número de acesso (*accession number*), tipo de sequência (*nt* ou *aa*), organismo, biblioteca, dentre outros (XIONG, 2006).

Visando o aperfeiçoamento do processo de cultivo da soja, no que diz respeito a fatores bióticos e abióticos que afetam a produção e produtividade nacional deste grão tais como seca, déficit hídrico, pragas e patógenos, no ano de 2007, o governo brasileiro iniciou o Consórcio Nacional para Estudos do Genoma da soja (GENOSOJA) (NASCIMENTO, 2009). O genoma da soja vem sendo amplamente estudado e investigado por vários grupos de estudos de diferentes países, incluindo o Brasil, por meio de estudos de genômica estrutural e funcional, transcriptômica e proteômica. Dentre os objetivos do GENOSOJA, estão o estudo da estrutura e função dos genes e proteínas, o perfil de expressão da soja quando submetida aos principais tipos de estresses biótico e abióticos que acometem a cultura do grão no Brasil e no mundo, tais como ferrugem asiática, nematoides, seca e salinidade. Todos os dados gerados a partir deste projeto serão depositados tanto em bancos de dados locais e de acesso restrito no site do GENOSOJA, como também em bancos de dados públicos como o GenBank (www.ncbi.nlm.gov/genbank/) (ABDELNOOR *et al.*, 2009).

A rede NordEST teve início em 2005, através da Rede Nordestina de Biotecnologia (ReNorBio), sob a coordenação da Universidade Federal de Pernambuco. Nesta rede, inicialmente, foram realizadas análises de genômica funcional, estrutural e comparativa no feijão-caupi (*V. unguiculata*), com o objetivo principal de identificar genes candidatos potencialmente úteis para fins de melhoramento nesta cultura. A rede integra 10 instituições e 11 laboratórios de várias regiões do Brasil. No âmbito desta rede foram realizados estudos de transcriptômica, incluindo o sequenciamento de etiquetas de sequências expressas (ESTs - *Expressed Sequence Tags*), Análise Serial da Expressão Gênica (HT-SuperSAGE) e de RNA-Seq. O banco NordEST inclui 12 bibliotecas de EST sob condições de estresse biótico (vírus do mosaico severo e potyvírus) e abiótico (seca e salinidade) incluindo mais de 500 milhões de transcritos sob diversas condições até 2010 (Benko-Iseppon *et al.*, 2010). Desta forma o banco disponibiliza informações transcripcionais de uma diversidade de genes em vários momentos de resposta vegetal frente às situações contrastantes aplicadas experimentalmente (Benko-Iseppon, 2009). O mapeamento físico dos cromossomos está em desenvolvimento, incluindo amostras de microssatélites, rDNA e BACs, o que torna viável a rápida transferência de dados entre culturas visando ao melhoramento por meio de estudos de genômica estrutural e funcional, transcriptômica e proteômica. Dentre os objetivos do GENOSOJA, estão o estudo da estrutura e função dos genes e proteínas, o perfil de expressão da soja quando submetida aos principais tipos de estresses biótico e abióticos que acometem a cultura do grão no Brasil e no mundo, tais como ferrugem asiática, nematoides, seca e salinidade. Todos os dados gerados a partir deste projeto serão depositados tanto em bancos de dados locais e de acesso restrito no site do

GENOSOJA, como também em bancos de dados públicos como o GenBank (www.ncbi.nlm.gov/genbank/) (ABDELNOOR *et al.*, 2009).

3 ESTRUTURA E FUNÇÃO DE FATORES DE TRANSCRIÇÃO DA FAMÍLIA NAC EM SOJA E FEIJÃO-CAUPI SOB DESIDRATAÇÃO RADICULAR: UMA ANÁLISE COMPARATIVA¹

SANTOS, R. F.²; BEZERRA-NETO, J. P.²; PANDOLFI, V.²; LIMA, M. O.²; BINNECK, E.³; BENKO-ISEPPON, A. M.²

RESUMO

Genes da família NAC (NAM, ATAF1/2 e CUC2) codificam fatores de transcrição (FTs) específicos de plantas que desempenham importantes funções tanto no desenvolvimento vegetal, quanto na regulação dos mecanismos de defesa contra estresses bióticos e abióticos. Dentre os estresses abióticos, a seca ocasiona reduções drásticas nos níveis de produtividade de culturas de interesse agrônomo como a soja e o feijão-caupi. Neste trabalho, sequências sondas de NAC de *Arabidopsis thaliana* e de quatro espécies da família Fabaceae, disponíveis em bancos de dados, foram utilizadas na identificação, caracterização estrutural e funcional de 50 e 42 FTs NAC dos transcriptomas da soja e do feijão-caupi, respectivamente, submetidos a estresse por desidratação radicular comparativamente a um controle não estressado. Além da anotação funcional dos genes candidatos e suas respectivas proteínas, foram realizadas análises da estrutura de domínios e motivos funcionais, predição de localização subcelular de estrutura secundária, análise fenética, modelagem tridimensional da estrutura proteica. Os resultados indicam significativa diversidade e diferenças na modulação da expressão desses fatores nas leguminosas analisadas, sendo potencialmente úteis para uso como marcadores moleculares ou em transgênese, visando melhor desempenho sob estresses bióticos e abióticos.

Palavras-chave: *Vigna unguiculata*. *Glycine max*. Bioinformática. Seca. Estresse abiótico.

¹ Artigo a ser submetido à revista Genetics and Molecular Biology (ISSN:1415-4757).

² Laboratório de Genética e Biotecnologia Vegetal, Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brasil.

³ Embrapa Soja. Rod. Carlos João Strass. Londrina PR, Brasil.

ABSTRACT

NAC genes (NAM, ATAF1/2 and CUC2) encode plant specific transcription factors (TFs) which perform important functions both on plant development and regulation of defense mechanisms against biotic and abiotic stresses. Among the types of abiotic stress, the drought causes drastic losses on productive levels of crops such as soybean and cowpea. In this study, NAC seed sequences of *Arabidopsis thaliana* and four Fabaceae species available public data banks were used to characterize structurally and functionally 50 and 42 NAC TFs in the soybean and cowpea transcriptomes, respectively, submitted to root dehydration as compared with non-stressed controls. Besides the functional annotation, analysis was carried out, including identification and characterization of conserved motifs and domains, subcellular localization and secondary structure prediction, phylogenetic analysis, tridimensional modelling of protein structure. The results indicate a significant diversity of these factors and differences in their expression modulation in the analyzed legumes, being potentially useful as molecular markers or in transgenesis, aiming at a better performance of the plants under biotic and abiotic stresses.

Keywords: *Vigna unguiculate*. *Glycine max*. Bioinformatics. Drought. Abiotic stress.

3.1 INTRODUÇÃO

Fatores de transcrição (FTs) são proteínas específicas que regulam a expressão gênica através da ligação de sequências específicas de DNA, na região regulatória dos genes-alvos, modulando a atividade da RNA polimerase para o início do processo transcricional.

Os membros da família NAC – incluindo NAM (*No apical meristem*), ATAF 1/2 (*Arabidopsis transcription activator factor 1/2*) e CUC2 (*Cup shaped cotyledon*) são específicos de plantas (SOUER *et al.*, 1996; AIDA *et al.*, 1997; ZHENG *et al.*, 2016). NAC FTs possuem o domínio do sítio de ligação ao DNA altamente conservado na região N-terminal e um domínio funcional variável na região C-terminal (ERNST *et al.*, 2004; OLSEN *et al.*, 2005) a qual pode atuar tanto como um ativador, quanto como um repressor transcricional (TRAN *et al.*, 2004; HU *et al.*, 2006; KIM *et al.*, 2007). Por se tratar de uma grande família genica, membros NAC FTs têm sido identificados em diversas plantas, a exemplo de *Populus*, arroz, *Arabidopsis* e soja (HU *et al.*, 2010; NURUZZAMAN *et al.*, 2010; PINHEIRO *et al.*, 2009). Em adição ao seu importante papel na tolerância das plantas a estresses bióticos e abióticos (NAKASHIMA *et al.*, 2012; XU *et al.*, 2014), NAC FTs atuam direta e indiretamente sobre o crescimento e desenvolvimento das plantas, incluindo o desenvolvimento embrionário, germinação da semente, desenvolvimento de raízes laterais, formação floral, resposta imunológica, divisão e crescimento celular, senescência e formação da parede secundária (OLSEN *et al.*, 2005; LIU *et al.*, 2014).

Dentre os estresses abióticos que acometem culturas como a soja (*Glycine max*) e o feijão-caupi (*Vigna unguiculata*), a seca figura como um dos mais impactantes sobre a produtividade, sendo responsável por enormes perdas nas safras anuais destas culturas (MANTRI *et al.*, 2012; DO VALE *et al.*, 2012). Segundo Bray (2004), perdas ocasionadas em decorrência da seca chegam a comprometer mais de 50% da produção mundial destes grãos, enquanto os prejuízos ocasionados pelos estresses bióticos, como nematoides e viroses, comprometem a produtividade destas culturas entre 10 % e 20 %. Estima-se que na África sejam perdidas anualmente cerca de 300 mil toneladas de grãos de feijão-caupi devido à seca (RAMALHO *et al.*, 2009).

Nesse contexto, uma abordagem mais ampla e promissora para a obtenção de plantas mais tolerantes aos diversos tipos de estresses bióticos e abióticos é o uso biotecnológico de fatores de transcrição (YAMAGUCHI-SHINOZAKI *et al.*, 2002). A superexpressão de FTs, promotores estresse-induzidos ou ativadores transcricionais de sequências multigênicas, cujos produtos conferem direta e indiretamente um aumento na tolerância ao estresse apresenta-se

como uma alternativa promissora para alcançar resultados satisfatórios na contenção dos danos ocasionados por tais fatores de estresse (KASUGA *et al.*, 2004; PELLEGRINESCHI *et al.*, 2004; BECK *et al.*, 2007).

O presente estudo teve como objetivo identificar e caracterizar membros da família NAC de soja e feijão-caupi, bem como avaliar seu comportamento a nível transcricional em resposta à seca, inferindo sobre sua diversidade estrutural e funcional, bem como sobre seu papel na modulação da resposta dessas leguminosas contra o referido estresse.

3.2 MATERIAIS E MÉTODOS

3.2.1 Identificação de transcritos NAC de soja e feijão-caupi

Sequências proteicas de NAC descritas em *Arabidopsis thaliana*, *Arachis hypogaea*, *Cicer arietinum*, *Medicago truncatula* e *Phaseolus vulgaris* (Genbank/NCBI) foram usadas como sondas (*seed sequences*) na busca, via tBLASTn, de membros da família NAC de Fabaceae disponível no GenBank/NCBI. As 10 sequências com melhor resultado de alinhamento (*cut-off* de e^{-05}) foram utilizadas na geração de um alinhamento via BLASTx (*cut-off* de e^{-05}) contra os bancos de dados de RNA-Seq de soja sob desidratação radicular (GENOSOJA, <http://bioinfo03.ibi.unicamp.br/soja/>) e do feijão-caupi (NordEST, <http://bioinfo03.ibi.unicamp.br/vigna/> e o banco do transcriptoma do feijão-caupi depositado no banco de dados da EMBRAPA Londrina, <http://bioinfo.cnpso.embrapa.br/caupi/index.php>) na identificação de transcritos similares aos membros da família NAC nestas espécies.

3.2.2 Determinação dos domínios e motivos conservados

De posse das sequências de RNA-Seq obtidas nos bancos de dados (GENOSOJA e NordEST), os candidatos a membros NAC foram analisados com o auxílio da ferramenta ORF-Finder (NCBI) para a identificação dos quadros abertos de leitura (ORFs) e as possíveis regiões codificantes das sequências e seus possíveis produtos proteicos. As sequências anotadas foram submetidas a um alinhamento via BLASTp para a identificação de proteínas ortólogas, bem como seus respectivos domínios conservados com o auxílio da ferramenta CD-Search do BLASTp, sendo os resultados confirmados através da ferramenta Batch CD-Search – NCBI (MARCHLER-BAUER *et al.*, 2015). Os motivos conservados foram identificados através de análises de múltiplos alinhamentos usando a ferramenta MEME (*Multiple Expectation*

Maximization for Motif Elicitation) (MEMEBAILEY & ELKAN, 1994; BAILEY *et al.*, 2009). Para estas análises, foram utilizados os seguintes parâmetros: número de repetições – “*any number of repetitions*”; número máximo de motivos – 20; tamanho dos motivos – de 6 a 200 aminoácidos. Tais parâmetros foram definidos para realizar uma análise mais aprofundada quanto à diversificação dos genes NAC tanto em soja, quanto em feijão-caupi. Também foram analisados e identificados os padrões de distribuição desses motivos nas sequências identificadas. Sequências sem domínio conservado, com domínio incompleto ou diferente do característico da família NAC foram descartadas.

3.2.3 Localização subcelular

A predição da localização subcelular dos membros NAC foi realizada com o auxílio da ferramenta ProtComp (Softberry Inc.) com o objetivo de determinar comparativamente as estruturas proteicas dos genes expressos nos transcriptomas analisados com as disponíveis na literatura para NAC FTs.

3.2.4 Predição da estrutura secundária e modelagem tridimensional

As sequências proteicas anotadas de soja e feijão-caupi foram submetidas a um alinhamento múltiplo usando Clustal Omega (SIEVERS *et al.*, 2011) a fim de avaliar o nível de conservação estrutural entre elas. Estes resultados foram visualizados no programa BOXSHADE (http://embnet.vital-it.ch/software/BOX_form.html).

A predição da estrutura secundária foi conduzida com o programa JPred4 (DROZDETSKIY *et al.*, 2015; <http://www.compbio.dundee.ac.uk/www-jpred/>). Para automatizar estas análises, o *script* bash (shell) foi adaptado para a execução da rotina de análises do programa em todas as sequências de soja e feijão-caupi. As sequências de cada uma das duas espécies, pertencentes às diferentes subfamílias de NAC, foram selecionadas para a modelagem tridimensional através do programa SWISS-MODEL (ARNOLD *et al.*, 2006; GUEX *et al.*, 2009; KIEFER *et al.*, 2009; BIASINI *et al.*, 2014).

3.2.5 Análise fenética da superfamília nac em soja e feijão-caupi

As sequências proteicas de soja e feijão-caupi foram alinhadas através do programa Clustal Omega (SIEVERS *et al.*, 2011), sendo analisadas com o auxílio do software MEGA6

(TAMURA *et al.*, 2013) com posterior geração da árvore fenética, construída pelo método *Neighbor-Joining* (SAITOU & NEI, 1987), com *bootstrap* fixado em 1.000 réplicas e reconstrução pelo método de Dayhoff (SCHWARZ & DAYHOFF, 1979). As sequências foram agrupadas de acordo com os subtipos de NAC descritos para soja (NAC 1- 6) e disponíveis no GenBank-NCBI.

3.2.6 Análise de expressão *in silico* baseada em dados de rna-seq

Com base nos dados de valores de expressão FPKM (*Fragments per kilobase of transcript per million fragments mapped*) disponíveis nos bancos de dados de RNA-Seq da soja e do feijão-caupi, foi gerado um perfil de expressão *in silico* (*heatmap*) dos transcritos de NAC em seus respectivos transcriptomas. Os dados foram clusterizados hierarquicamente com o auxílio do programa Cluster 3.0 (EISEN *et al.*, 1998) e visualizados com o auxílio do programa Java TreeView (SALDANHA, 2004).

3.2.7 Desenho de *primers*

Com base nos níveis de expressão (obtidos pelo *fold-change -FC*) identificados a partir dos bancos de dados de RNA-Seq da soja e do feijão-caupi, foram selecionadas sequências para desenho de um conjunto de *primers* visando a posterior amplificação de genes codificantes dos NAC FTs para cada uma das espécies. Para isso, foi usado o programa Primer3 Plus (UNTERGASSER *et al.* 2007), utilizando-se como padrões: tamanho de 20pb, temperatura de anelamento entre 59.9 e 60.2°C e produto de amplificação variando entre 150 e 226 pb. Estes *primers* serão utilizados para a validação da expressão via PCR quantitativa em tempo real (*Real Time quantitative PCR*, RT-qPCR).

3.3 RESULTADOS E DISCUSSÃO

3.3.1 Identificação de transcritos nac de soja e feijão-caupi

As sequências-sonda de NAC FTs obtidas no GenBank foram anotadas, via tBLASTn contra o banco de Fabaceae do próprio NCBI visando à obtenção de sequências de seus genes codificantes. A partir de cada sonda proteica inicial de cada organismo de referência, foram anotadas as 10 sequências nucleotídicas com os melhores resultados, usando-se o *e-value* como

referência para a classificação, com um ponto de corte (*cut-off*) o valor e^{-05} , obtendo-se 170 sequências nucleotídicas (Tabela 1).

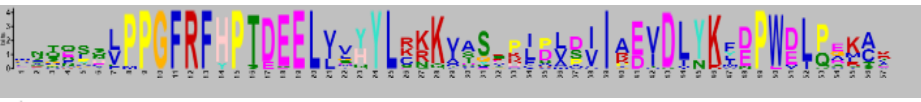
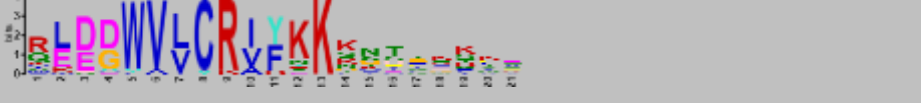
Tabela 1 — Número de *seed sequences* proteicas dos fatores de transcrição da família NAC (NAC FTs) anotadas, com base na literatura, para cada organismo de referência depositadas no Genbank – NCBI e quantidade de sequências de transcritos obtidos via tBLASTn contra o banco de Fabaceae do GenBank – NCBI ($cut-off \leq e^{-05}$)

Espécie	No. sequências NAC	No. transcritos (tBLASTn)
<i>Arabidopsis thaliana</i>	05	50
<i>Arachis hypogaea</i>	03	30
<i>Cicer arietinum</i>	01	10
<i>Medicago truncatula</i>	05	50
<i>Phaseolus vulgaris</i>	03	30
TOTAL	17	170

Fonte: o autor (2017).

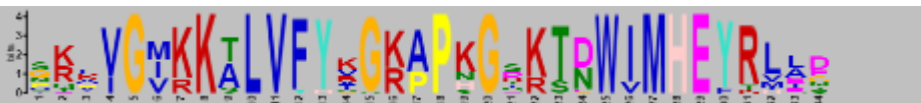
Em seguida, todas as 170 sequências de genes codificantes anotadas foram compiladas em um arquivo *MULTIFASTA*, sendo posteriormente analisadas contra os bancos de dados de RNA-Seq da soja (GENOSOJA) e do feijão-caupi (NordEST). Com o auxílio da ferramenta ORF-Finder, foram identificadas as ORFs e as possíveis regiões funcionais (codificantes) dos genes de interesse, bem como seus respectivos produtos proteicos. Durante essa etapa, não foi possível identificar, em algumas sequências, ORFs, possivelmente por se tratarem de sequências incompletas, não funcionais, ou pseudogenes, sendo tais sequências descartadas das etapas de análises posteriores. As sequências proteicas referentes aos produtos dos genes analisados nessa etapa foram submetidas a uma análise via BLASTp para a obtenção de proteínas homólogas, bem como para a identificação, via CD-Search, dos domínios conservados (Tabelas 2 e 3).

Tabela 2 — Relação dos motivos conservados mais abundantes identificados nas sequências de soja com o auxílio da ferramenta MEME. Estes motivos ocorrem em mais de 88% das sequências analisadas

Motivo	Logo	Ocorrência
01		50/50
02		44/50
03		50/50
04		49/50

Fonte: o autor (2017).

Tabela 3 — Relação dos motivos conservados mais abundantes identificados nas sequências de feijão-caupi com o auxílio da ferramenta MEME. Estes motivos ocorrem em mais de 83% das sequências analisadas

Motivo	Logo	Ocorrência
01		42/42
02		42/42
03		42/42
04		35/42
05		41/42

Fonte: o autor (2017).

3.3.2 Determinação dos domínios e motivos conservados

As sequências das proteínas homólogas foram submetidas a uma nova análise de busca e identificação de domínio conservado, via Batch CD-Search, para que os resultados fossem confirmados e, nos casos de domínios conservados desconhecidos, estes fossem devidamente identificados.

Como a família gênica NAC tem seu nome derivado dos três primeiros genes contendo esse domínio (*NAM*, *No apical meristem*, *ATAF1/2* e *CUC2*, *Cup-shaped cotyledon*), é comum alguns membros dessa família apresentarem o domínio “*NAM superfamily*” como domínio conservado. Estas, por sua vez, são proteínas vegetais específicas que estão associadas ao desenvolvimento estrutural das plantas e, como é característico da família NAC, também atuam nas vias de resposta a estresses bióticos e abióticos, bem como na regulação dos demais genes atuantes nessas vias (SOUER *et al.*, 1996; AINDA *et al.*, 1997; HU *et al.*, 2010). Dessa forma, todas as sequências que apresentaram domínio conservado inexistente ou diferente do domínio característico, foram descartadas das etapas de análises posteriores (Tabelas 2 e 3).

O domínio NAC pode ser subdividido em subdomínios cuja quantidade pode variar de espécie para espécie. Cada subdomínio pode ser identificado e classificado de acordo com a composição específica de aminoácidos de seus respectivos motivos. Kikuchi *et al.* (2000) identificaram 5 (A – E) subdomínios NAC em arroz e afirmou que em geral, os subdomínios são ricos em aminoácidos básicos (R, K e H); porém, o que difere um subdomínio dos outros é a distribuição de aminoácidos positivos e negativos e suas relativas abundâncias nos motivos.

Para a identificação dos domínios NAC, bem como dos motivos e padrões conservados que os compõem, foram realizadas análises de predição de motivos, através da ferramenta MEME (*Multiple Expectation Maximization for Motif Elicitation*), que identifica motivos contínuos (sem *gaps* ou intervalos) em regiões conservadas de sequências de DNA, RNA e proteína.

Todas as sequências analisadas, tanto para soja, quanto para feijão-caupi, apresentaram índices de conservação elevados em motivos, atendendo aos padrões de distribuição dos subdomínios, bem como em relação à abundância relativa de aminoácidos específicos nestes. Foram identificados 20 motivos conservados distintos em ambas as espécies (Figuras 1 e 2).

Em soja, os motivos mais abundantes foram 1, 2, 3 e 4 (Tabela 2), enquanto que em feijão-caupi, foram os motivos 1, 2, 3, 4 e 5 (Tabela 3). Resultados semelhantes foram encontrados em *A. thaliana* (ERNST *et al.*, 2004) e em *Eucalyptus grandis* (HUSSEY, *et al.*, 2015). De acordo com estes autores, por se tratar de uma família gênica bastante versátil, espera-se que sejam encontrados “motivos gerais” como estes. Acredita-se que tais motivos sejam os responsáveis por conferir esse caráter generalista aos genes da família NAC. Estes motivos geralmente encontram-se situados nas regiões próximas à extremidade N-terminal, apresentando tamanho e composição variável, contudo, apresentando certa predominância geral de aminoácidos básicos (R, K e H). Os demais motivos não coincidiram com os motivos

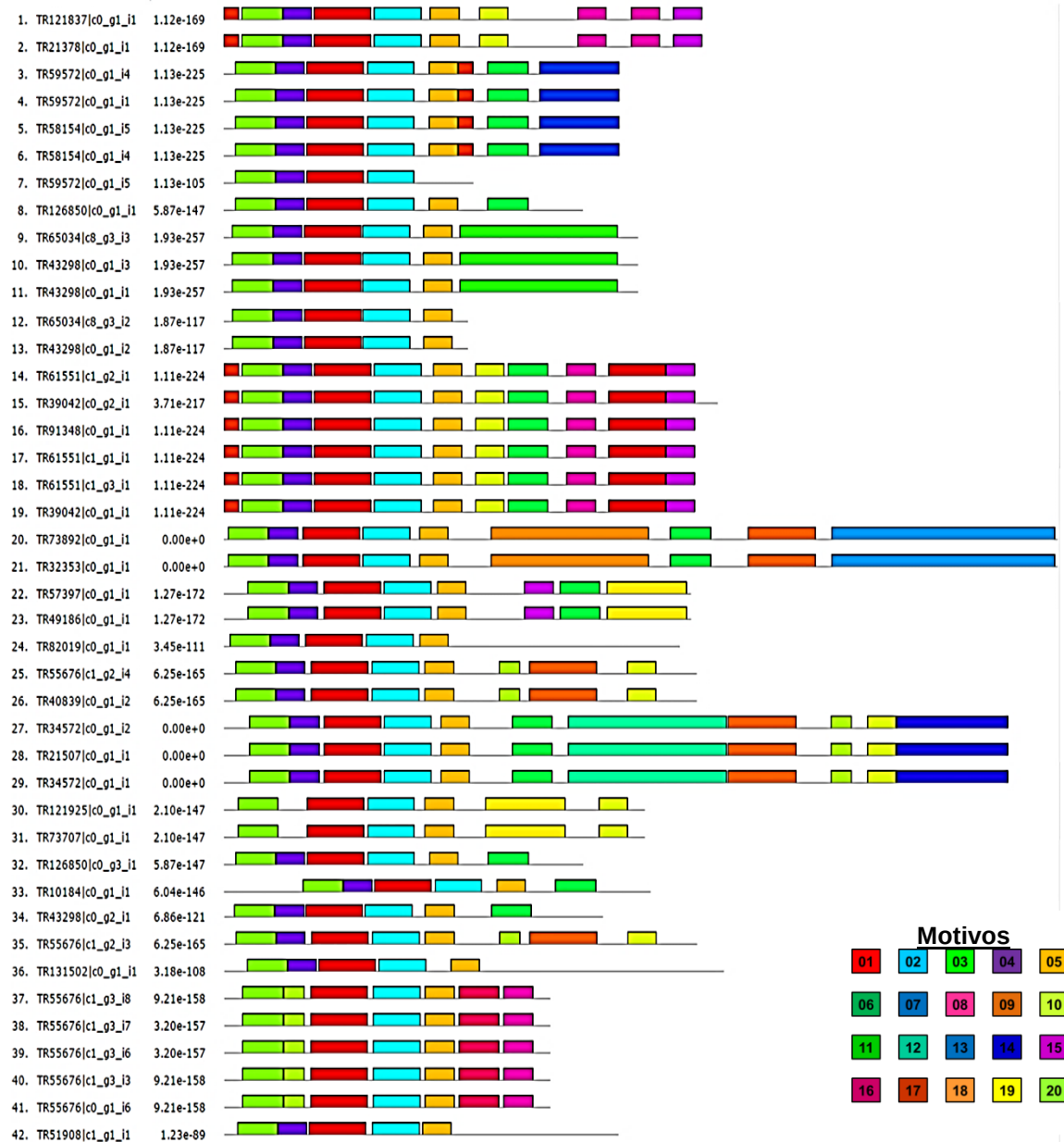
descritos em arroz, *A. thaliana* e *Populus trichocarpa* (OOKA *et al.*, 2003; FANG *et al.*, 2008; HU *et al.*, 2010; JENSEN *et al.*, 2010).

Sequências específicas (menos abundantes ou restritas a uma sequência ou grupo de sequências), como o motivo 12 em feijão-caupi e o motivo 17 em soja, podem ser motivos espécie-específicos ou desempenhar funções específicas para aquele determinado gene ou grupo de genes (ERNST *et al.*, 2004; HUSSEY *et al.*, 2015). Tal fato pode ser verificado nas análises fenéticas, onde as sequências analisadas foram agrupadas em subfamílias, cujos motivos que compõem estes agrupamentos apresentam alto grau de similaridade, sugerindo, inclusive, a existência de motivos homólogos entre feijão-caupi e soja.

3.3.3 Localização subcelular

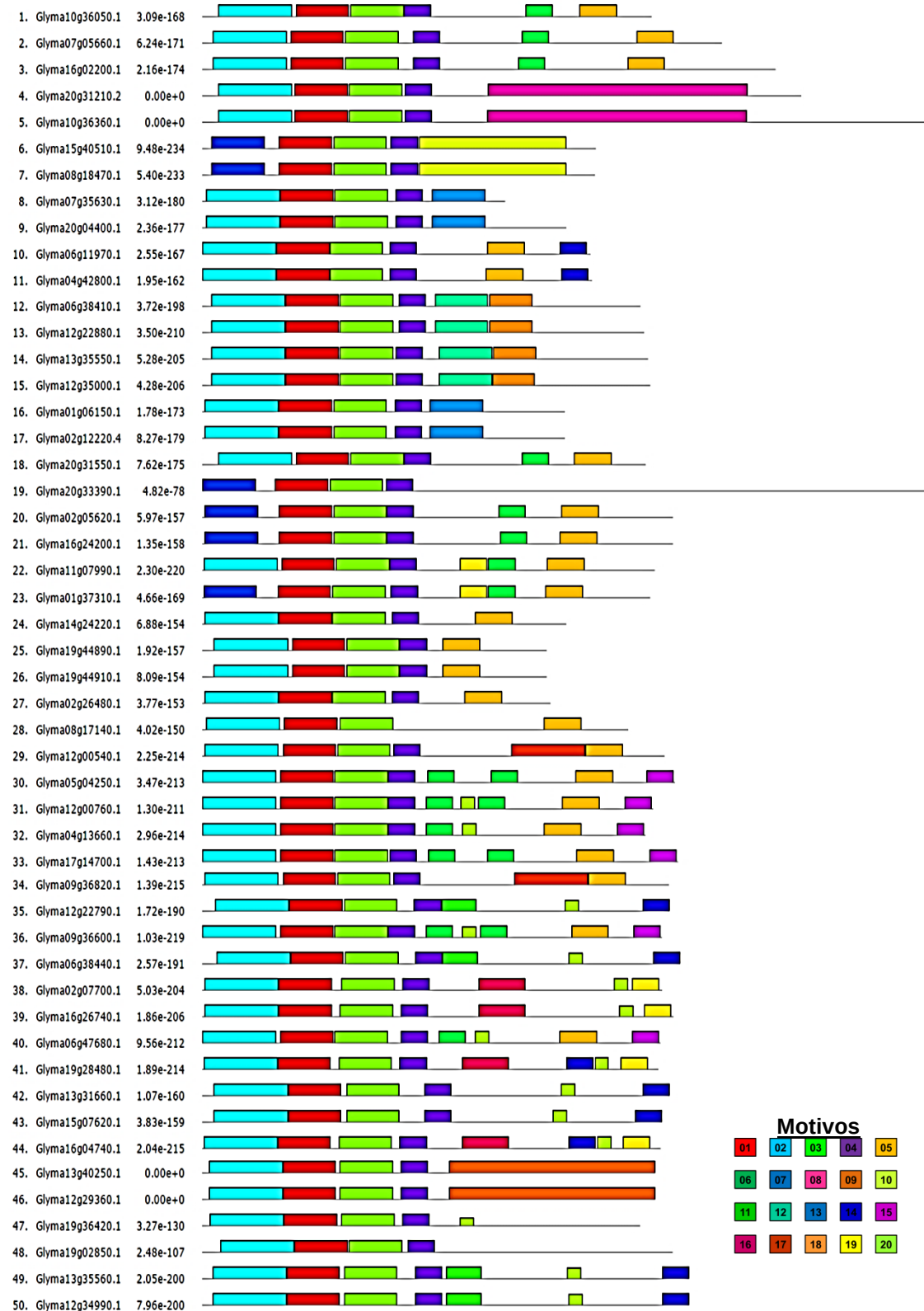
Com o objetivo de enriquecer a análise e confrontar os dados obtidos com aqueles disponíveis na literatura, foram realizadas análises *in silico* de predição da localização subcelular destes FTs com a ferramenta ProtComp (Tabelas 4 e 5). Por se tratar de fatores de transcrição, presume-se que sejam localizados no núcleo, embora relatos apontem para alguns representantes da família NAC que não apresentam os padrões habituais de sinais de localização nuclear (TAOKA *et al.*, 2004; JIN *et al.*, 2013). Os níveis de confiabilidade são indicados numa escala de 0 a 10, onde quanto mais próximo de 10, mais confiáveis são as predições. Com base nos resultados, 11,9 % das sequências de feijão-caupi e 6 % das sequências de soja são excretadas. Contudo, por se tratar de Fatores de Transcrição, espera-se que estes sejam encontrados no núcleo, com exceção daqueles FTs que apresentam como característica uma localização diferente desta.

Figura 1 — Representação esquemática dos motivos conservados das sequências proteicas de NAC do feijão-caupi (*V. unguiculata*) obtidos através da ferramenta MEME. Cada motivo é representado por uma cor e um número. As linhas pretas abaixo das caixas coloridas representam as regiões não conservadas das sequências



Fonte: o autor (2017).

Figura 2 — Representação esquemática dos motivos conservados das sequências proteicas de NAC da soja (*G. max*) obtidos através da ferramenta MEME. Cada motivo é representado por uma cor e um número. As linhas pretas abaixo das caixas coloridas representam as regiões não conservadas das sequências



Fonte: o autor (2017).

Tabela 4 — Relação dos transcritos codificadores para NAC encontrados no genoma da soja (*Glycine max*). Dados adicionais das sequências obtidos no NCBI após BLASTp e Batch CD-Search juntamente com os dados de predição da localização subcelular obtidas com o auxílio da ferramenta ProtComp (SoftBerry Inc.)

Gene	BLASTp			Batch CD-Search		ProtComp	
	Acesso NCBI	Tamanho (aa)	e-value	Melhor alinhamento	Domínio	Status	Localização Score
Glyma10g36050.1	XP_006589457.1	346	0	<i>Glycine max</i>	NAM	Completo	Nuclear 6.9
Glyma07g05660.1	XP_003529874.1	400	0	<i>Glycine max</i>	NAM	Completo	Nuclear 8.9
Glyma16g02200.1	KHN15891.1	441	0	<i>Glycine soja</i>	NAM	Completo	Nuclear 8.8
Glyma20g31210.2	KRG91746.1	461	0	<i>Glycine max</i>	NAM	Completo	Extracelular (Excretado) 3
Glyma10g36360.1	NP_001274375.1	560	0	<i>Glycine max</i>	NAM	Completo	Extracelular (Excretado) 2.5
Glyma15g40510.1	NP_001235901.1	303	0	<i>Glycine max</i>	NAM	Completo	Nuclear 8.8
Glyma08g18470.1	KHN26800.1	302	0	<i>Glycine max</i>	NAM	Completo	Nuclear 8.8
Glyma07g35630.1	KHN28140.1	233	0	<i>Glycine soja</i>	NAM	Completo	Nuclear 9
Glyma20g04400.1	XP_003556789.1	280	2.00E-180	<i>Glycine max</i>	NAM	Completo	Nuclear 8.9
Glyma06g11970.1	ALA09235.1	299	0	<i>Glycine max</i>	NAM	Completo	Nuclear 8.9
Glyma04g42800.1	NP_001240958.1	300	0	<i>Glycine max</i>	NAM	Completo	Nuclear 9
Glyma06g38410.1	AHB63576.1	337	0	<i>Glycine max</i>	RhaD_aldol-ADH	Completo	Nuclear 8.9
Glyma12g22880.1	NP_001304273.1	340	0	<i>Glycine max</i>	NAM	Completo	Nuclear 8.8

Glyma13g35550.1	AHB63579.1	343	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.9
Glyma12g35000.1	ALA09241.1	345	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma01g06150.1	NP_001236855.1	279	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.9
Glyma02g12220.4	KHN14494.1	279	1.00E-108	<i>Glycine soja</i>	NAM	Completo	Nuclear	9
Glyma20g31550.1	XP_006606215.1	341	0	<i>Glycine max</i>	NAM	Completo	Nuclear	6.6
Glyma20g33390.1	XP_003556302.1	604	0	<i>Glycine max</i>	NAM	Completo	Extracelular (Excretado)	2.4
Glyma02g05620.1	XP_003520268.1	362	0	<i>Glycine max</i>	NAM	Incompleto	Nuclear	7.3
Glyma16g24200.1	XP_003547947.1	362	0	<i>Glycine max</i>	NAM	Completo	Nuclear	7.6
Glyma11g07990.1	XP_006590718.1	348	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.8
Glyma01g37310.1	KHN36047.1	345	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.8
Glyma14g24220.1	ALA09243.1	280	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma19g44890.1	XP_003553799.1	265	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma19g44910.1	KRG97237.1	265	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma02g26480.1	AGO14646.1	268	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma08g17140.1	XP_003531470.1	328	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
Glyma12g00540.1	XP_003540565.1	356	0	<i>Glycine max</i>	NAM	Ausente	Nuclear	8.8
Glyma05g04250.1	XP_006579623.1	364	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8
Glyma12g00760.1	XP_006591958.1	347	0	<i>Glycine max</i>	NAM	Completo	Nuclear	7.4
Glyma04g13660.1	XP_003523884.2	341	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.7

Glyma17g14700.1	XP_006600826.1	366	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
Glyma09g36820.1	XP_003533525.1	359	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.8
Glyma12g22790.1	KHN48759.1	360	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma09g36600.1	XP_006587710.1	354	0	<i>Glycine soja</i>	NAM	Completo	Nuclear	7.3
Glyma06g38440.1	NP_001274379.1	368	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
Glyma02g07700.1	XP_003518189.1	354	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
Glyma16g26740.1	XP_003548053.1	363	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.9
Glyma06g47680.1	KHN42768.1	353	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.7
Glyma19g28480.1	NP_001236142.1	351	7.00E-144	<i>Glycine soja</i>	NAM	Ausente	Nuclear	8.9
Glyma13g31660.1	KHN08560.1	360	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
Glyma15g07620.1	XP_003547137.1	354	0	<i>Glycine soja</i>	NAM	Completo	Nuclear	8.9
Glyma16g04740.1	KHN46799.1	353	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.9
Glyma13g40250.1	XP_003541998.1	349	3.00E-139	<i>Glycine soja</i>	NAM	Completo	Nuclear	8.8
Glyma12g29360.1	XP_003539409.1	349	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.8
Glyma19g36420.1	XP_003553546.2	337	1.00E-167	<i>Glycine max</i>	NAM	Incompleto	Nuclear	8.7
Glyma19g02850.1	XP_003554895.2	362	0	<i>Glycine max</i>	NAM	Completo	Nuclear	9
Glyma13g35560.1	XP_003543257.1	375	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.9
Glyma12g34990.1	XP_003540450.1	375	0	<i>Glycine max</i>	NAM	Completo	Nuclear	8.9

Tabela 5 — Relação dos transcritos codificadores para NAC encontrados no genoma do feijão-caupi. Dados adicionais das sequências obtidos no NCBI após BLASTp e Batch CD-Search juntamente com os dados de predição da localização subcelular obtidas com o auxílio da ferramenta ProtComp (SoftBerry Inc.)

Gene	BLASTp				Batch CD-Search		ProtComp	
	Acesso NCBI	Tamanho (aa)	e-value	Melhor alinhamento	Domínio	Status	Localização	Score
TR121837 c0_g1_i1	XP_017425691.1	341	0	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.9
TR21378 c0_g1_i1	XP_017425691.1	341	0	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.9
TR59572 c0_g1_i4	XP_014490520.1	282	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR59572 c0_g1_i1	XP_014490520.1	282	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR58154 c0_g1_i5	XP_014490520.1	282	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR58154 c0_g1_i4	XP_014490520.1	282	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR59572 c0_g1_i5	XP_014618210.1	178	1.00E-107	<i>Glycine max</i>	NAM	Completo	Nuclear	9
TR126850 c0_g1_i1	XP_017416154.1	256	4.00E-82	<i>Vigna angularis</i>	Ausente	Ausente	Nuclear	9.1

TR65034 c8_g3_i3	XP_014501798.1	295	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	9.1
TR43298 c0_g1_i3	XP_014501798.1	295	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	9.1
TR43298 c0_g1_i1	XP_014501798.1	295	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	9.1
TR65034 c8_g3_i2	ACD39367.1	174	2.00E-102	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
TR43298 c0_g1_i2	ACD39367.1	174	2.00E-102	<i>Glycine max</i>	NAM	Completo	Nuclear	9.1
TR61551 c1_g2_i1	XP_017433508.1	336	0	<i>Vigna</i> <i>angularis</i>	NAM	Completo	Nuclear	8.8
TR39042 c0_g2_i1	BAT89593.1	352	4.00E-129	<i>Vigna</i> <i>angularis</i>	NAM	Completo	Nuclear	8.8
TR91348 c0_g1_i1	XP_017433508.1	336	4.00E-87	<i>Vigna</i> <i>angularis</i>	Ausente	Ausente	Nuclear	8.8
TR61551 c1_g1_i1	XP_017433508.1	336	3.00E-106	<i>Vigna</i> <i>angularis</i>	NAM	Completo	Nuclear	8.8
TR61551 c1_g3_i1	XP_017433508.1	336	0	<i>Vigna</i> <i>angularis</i>	NAM	Completo	Nuclear	8.8

TR39042 c0_g1_i1	XP_017433508.1	336	3.00E-106	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.8
TR73892 c0_g1_i1	XP_014514424.1	594	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Extracelular (Excretado)	2.4
TR32353 c0_g1_i1	XP_014514424.1	594	0	<i>Vigna radiata</i> var. <i>radiata</i>	Ausente	Ausente	Extracelular (Excretado)	2.4
TR57397 c0_g1_i1	XP_017413977.1	333	0	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.5
TR49186 c0_g1_i1	XP_017413977.1	333	0	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.5
TR82019 c0_g1_i1	XP_017430339.1	325	1.00E-110	<i>Vigna angularis</i>	Ausente	Ausente	Nuclear	8.8
TR55676 c1_g2_i4	XP_014520143.1	337	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.8
TR40839 c0_g1_i2	XP_014520143.1	337	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.8
TR34572 c0_g1_i2	XP_017415221.1	559	0	<i>Vigna angularis</i>	NAM	Completo	Extracelular (Excretado)	2.4
TR21507 c0_g1_i1	XP_017415221.1	559	0	<i>Vigna angularis</i>	NAM	Completo	Extracelular (Excretado)	2.4

TR34572 c0_g1_i1	XP_017415221.1	559	0	<i>Vigna angularis</i>	NAM	Completo	Extracelular (Excretado)	2.4
TR121925 c0_g1_i1	XP_017423259.1	300	0	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.8
TR73707 c0_g1_i1	XP_017423259.1	300	0	<i>Vigna angularis</i>	NAM	Completo	Nuclear	8.8
TR126850 c0_g3_i1	XP_017416154.1	256	2.00E-54	<i>Vigna angularis</i>	Ausente	Ausente	Nuclear	9.1
TR10184 c0_g1_i1	XP_014506891.1	304	4.00E-67	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR43298 c0_g2_i1	XP_014503670.1	270	2.00E-157	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	9
TR55676 c1_g2_i3	XP_014520143.1	337	4.00E-127	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Incompleto	Nuclear	8.8
TR131502 c0_g1_i1	XP_007133059.1	356	7.00E-148	<i>Phaseolus vulgaris</i>	Ausente	Ausente	Nuclear	9
TR55676 c1_g3_i8	XP_014496413.1	232	1.00E-149	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR55676 c1_g3_i7	XP_014498272.1	232	1.00E-164	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9

TR55676 c1_g3_i6	XP_014498272.1	232	8.00E-149	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR55676 c1_g3_i3	XP_014496413.1	232	2.00E-144	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR55676 c0_g1_i6	XP_014496413.1	232	1.00E-149	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.9
TR51908 c1_g1_i1	XP_014495173.1	281	0	<i>Vigna radiata</i> var. <i>radiata</i>	NAM	Completo	Nuclear	8.6

Fonte: o autor (2017).

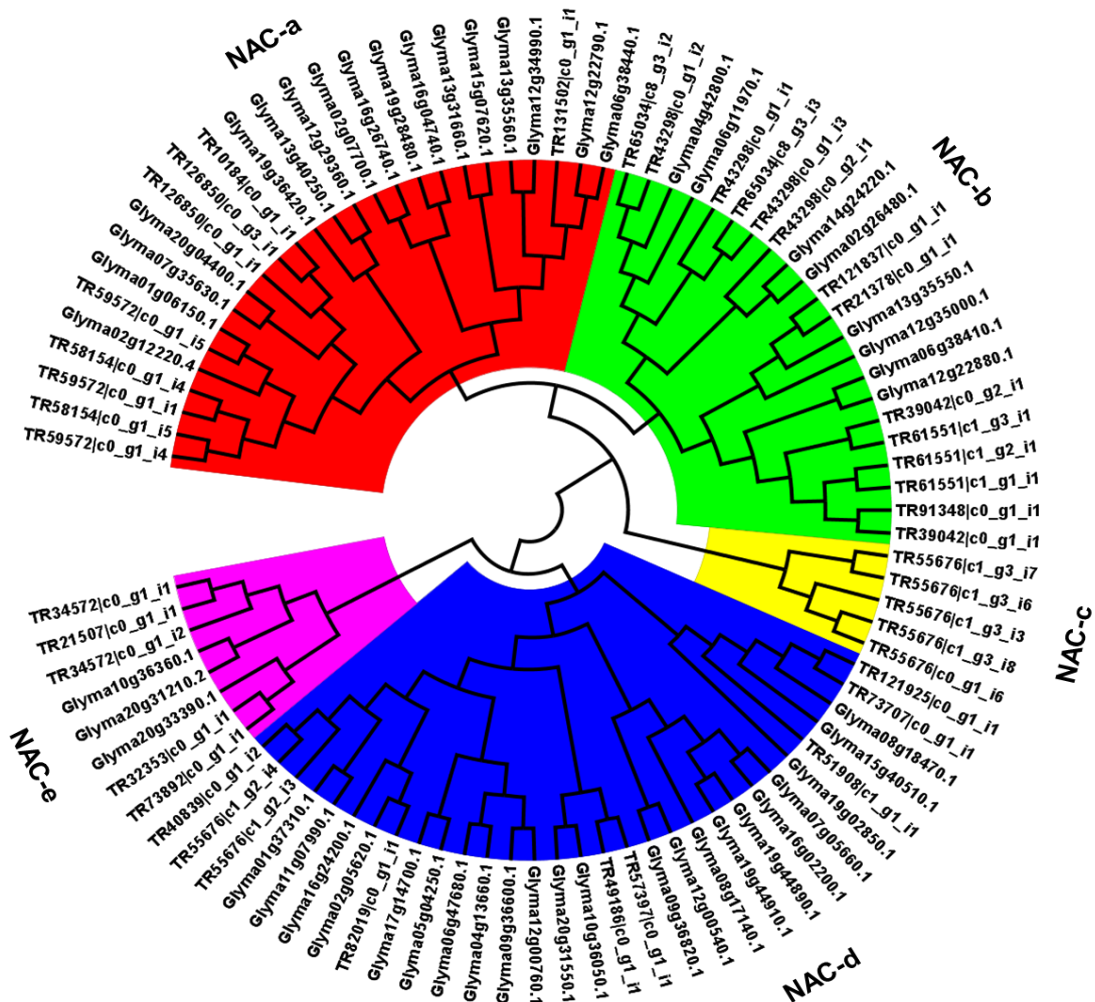
3.3.4 Análise fenética da superfamília NAC em soja e feijão-caupi

Os conjuntos de sequências de cada uma das duas espécies analisadas neste estudo foram submetidos a um alinhamento múltiplo via Clustal Omega com o objetivo de se analisar os níveis de conservação estrutural dentro de cada um destes conjuntos. Como era de se esperar, as sequências proteicas de soja e feijão-caupi apresentaram níveis elevados de conservação estrutural entre si. Adicionalmente, estas sequências foram utilizadas para a construção da árvore fenética (Figura 3). Estes resultados se assemelham aos resultados obtidos em *A. thaliana* e *E. grandis* (ERNST *et al.*, 2004; HUSSEY *et al.*, 2015).

De acordo com os resultados da análise fenética (Figura 3), foi possível verificar o agrupamento de algumas sequências das cinco subfamílias NAC (a – e) específicas com base na similaridade estrutural entre elas. Foi possível observar que na subfamília NAC-c não estavam agrupadas as sequências de NAC da soja (apenas sequências oriundas do transcriptoma do feijão-caupi), sugerindo que elas foram perdidas em soja ou adquiridas por feijão-caupi nas linhagens após os eventos de divergência desde o último ancestral comum de ambas. Como esperado, as sequências mais próximas (similares) que compõem as subfamílias observadas na árvore fenética apresentaram motivos conservados cujas composições foram similares dentro de cada subfamília, sugerindo similaridades funcionais entre as sequências proteicas que as compõem (Figuras 5, 6, 7, 8 e 9).

Com base nestes resultados e nos resultados do alinhamento múltiplo, uma sequência de cada uma das subfamílias de soja e de feijão-caupi foi selecionada para serem submetidas a um novo alinhamento múltiplo, via Clustal Omega, em conjunto com outras sequências proteicas de NAC depositadas no Genbank para outras espécies vegetais de referência como *A. thaliana*, *Arachis hypogaea*, *Medicago truncatula* e *Phaseolus vulgaris* (Figura 9), a fim de verificar se os padrões de conservação estrutural também eram compartilhados com as sequências destes organismos. De acordo com Hu *et al.* (2010), a similaridade estrutural e os motivos conservados dos genes NAC podem dar maior suporte para análises filogenéticas desta família. Entretanto, as diferenças da organização estrutural e da composição dos motivos conservados entre as diferentes subfamílias são forte indicativo de que esta família gênica é funcionalmente diversificada.

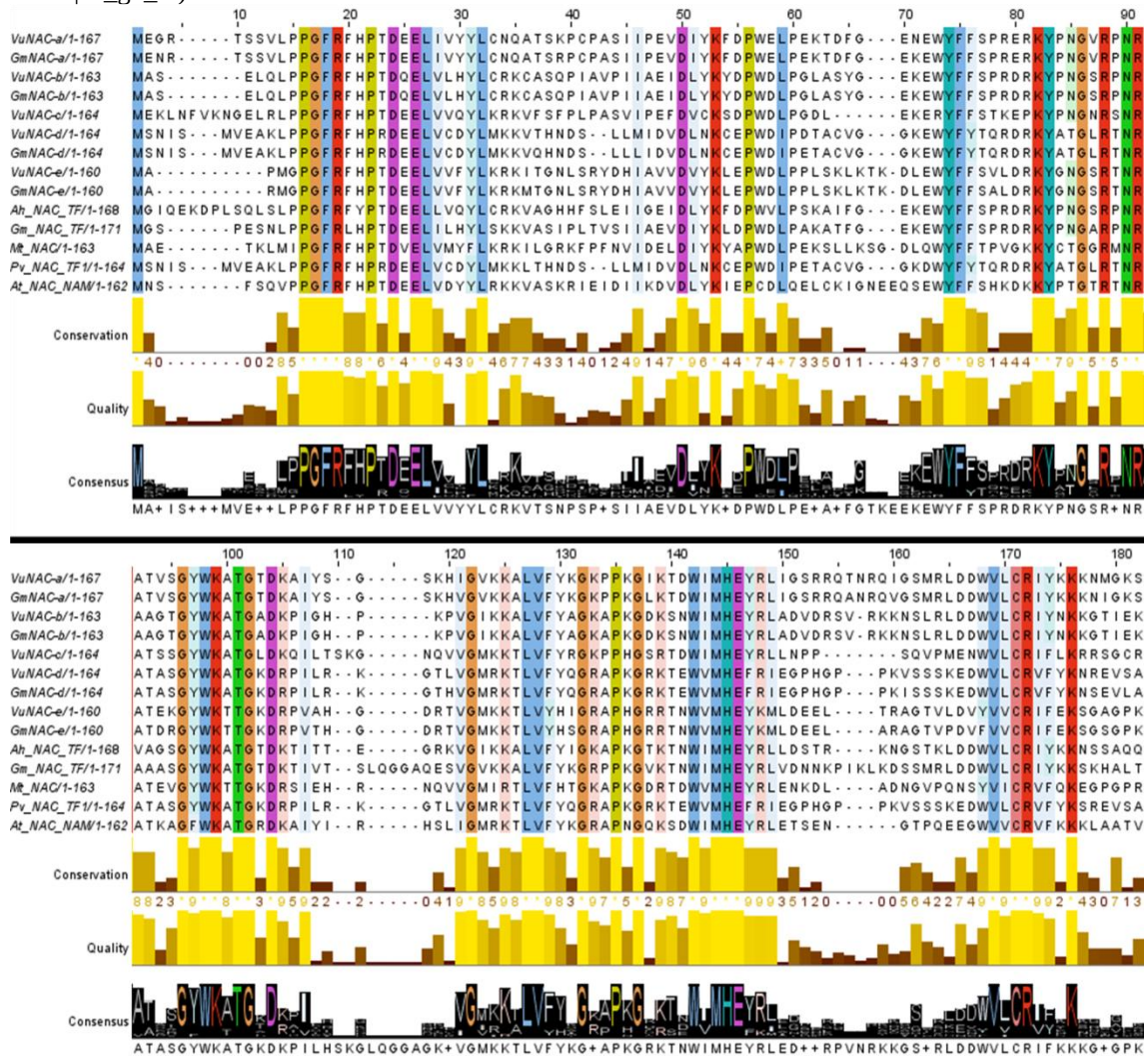
Figura 3 — As seqüências proteicas (*full-length*) de soja (Glyma = 50 seqüências) e feijão-caupi (TR = 42) foram alinhadas no Clustal Omega e a árvore fenética foi construída no programa MEGA6 através do método de Neighbor-Joining (NJ) com valor de bootstrap fixado em 1000. Cada subfamília NAC (a – e) está evidenciada por caixas coloridas: Vermelho – NAC-a; Verde – NAC-b; Amarelo – NAC-c; Azul – NAC-d; Lilás – NAC-e



Fonte: o autor (2017).

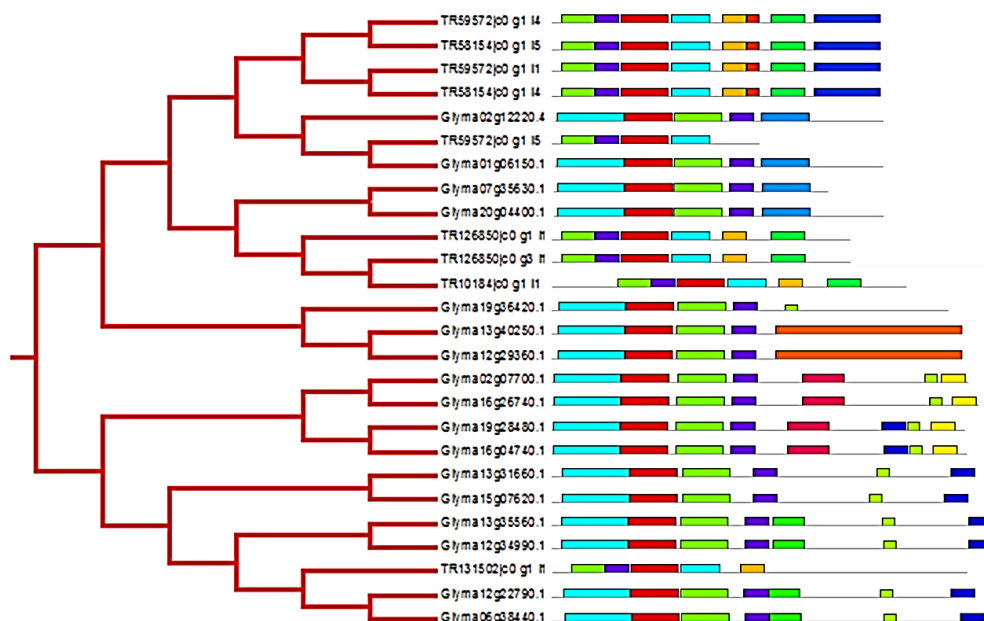
Contudo, o significado biológico específico da maioria dos motivos putativos identificados ainda não foi elucidado pelo fato de não possuírem homólogos quando analisados contra bancos de dados como Pfam e PDB.

Figura 4 — Edição do alinhamento múltiplo de algumas das sequências proteicas de NAC de soja (Gm), feijão-caupi (Vu), Medicago (Mt), feijão comum (Pv), amendoim (Ah) e Arabidopsis (At). Os resíduos vermelhos são comuns a no mínimo metade das sequências. Os resíduos azuis são quimicamente similares em mais da metade das sequências. O alinhamento corresponde à região com maiores níveis de conservação/similaridade entre as sequências analisadas. Mt_NAC (XP_013461558.1); Ah_NAC_TF (ADM94307.1); At_NAC_NAM (AEE33995.1); Pv_NAC_TF1 (AAK84883.1); GmNAC-a (Glyma02g12220.4); GmNAC-b (Glyma04g42800.1); GmNAC-d (Glyma08g18470.1); GmNAC-e (Glyma20g33390.1); VuNAC-a (TR59572|c0_g1_i4); VuNAC-b (TR65034|c8_g3_i2); VuNAC-c (TR55676|c1_g3_i7); VuNAC-d (TR121925|c0_g1_i1); VuNAC-e (TR73892|c0_g1_i1)



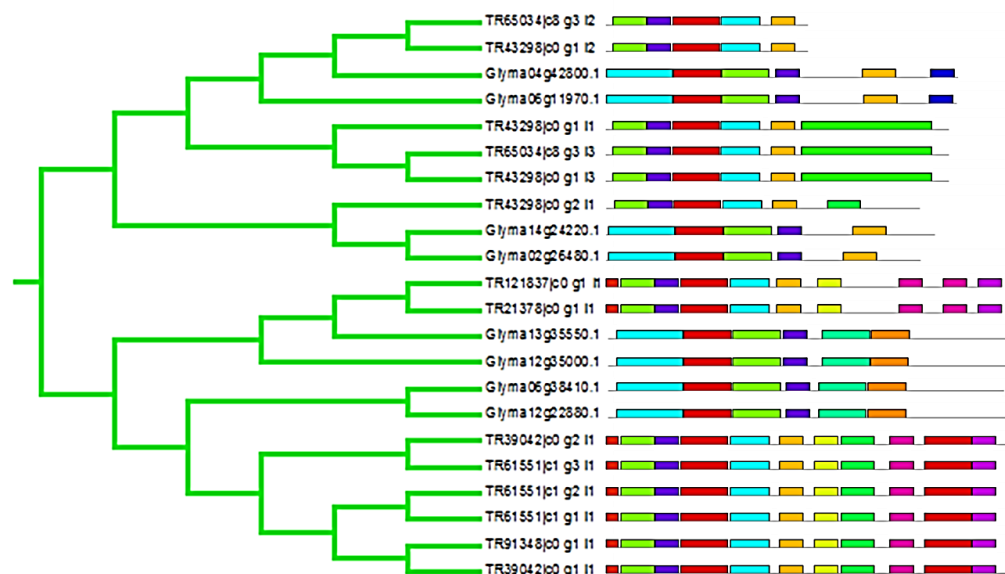
Fonte: o autor (2017).

Figura 5 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-a demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME



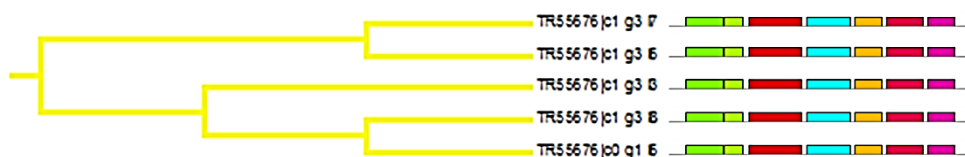
Fonte: o autor (2017).

Figura 6 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-b demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME



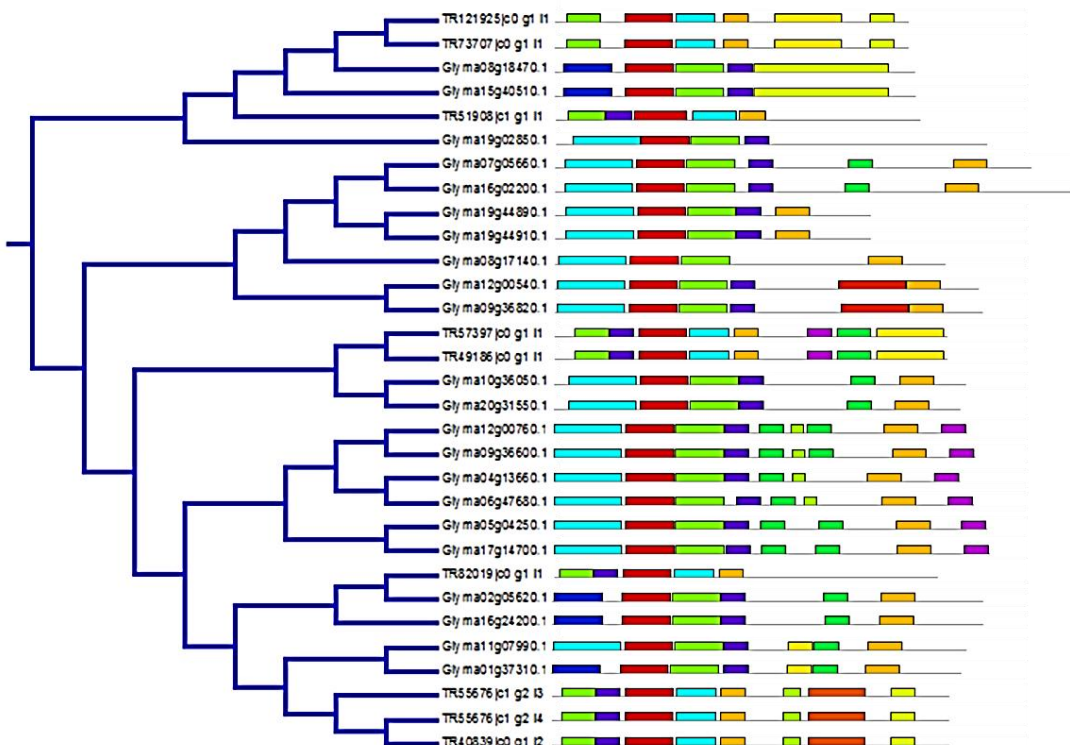
Fonte: o autor (2017).

Figura 7 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-c demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de feijão-caupi (TR) com o auxílio da ferramenta MEME



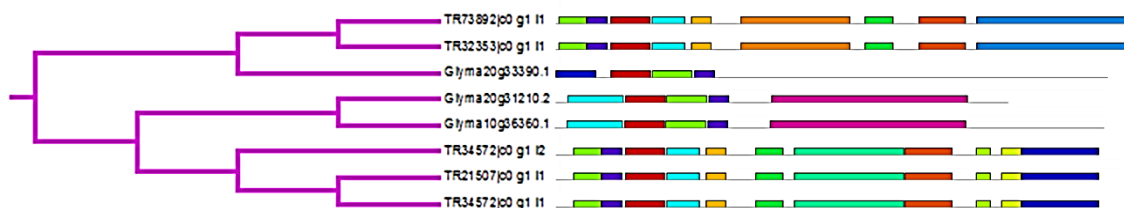
Fonte: o autor (2017).

Figura 8 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-b demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME



Fonte: o autor (2017).

Figura 9 — Representação esquemática da subseção da árvore fenética correspondente à subfamília NAC-b demonstrando a similaridade da composição dos motivos conservados identificados nas sequências de soja (Glyma) e feijão-caupi (TR) com o auxílio da ferramenta MEME



Fonte: o autor (2017).

3.3.5 Predição da estrutura secundária e modelagem tridimensional

Com o auxílio do programa JPred4 foi realizada a predição da estrutura secundária das sequências de soja e feijão-caupi. Para tais análises, como explicado na metodologia, foi utilizado o script `jpredapi` do pacote `jpredMassSubmitSchedule`. Objetivando a automatização desta etapa, um *script* bash (shell) foi escrito e as análises foram realizadas localmente, via comando de linha. O conteúdo observado da estrutura secundária das sequências analisadas está de acordo com os achados descritos em estudos como o de Ernst *et al.* (2004) e Greve *et al.* (2003).

De um modo geral, a estrutura secundária do domínio NAC é composta por uma folha- β bastante retorcida que se encaixa com uma α -hélice N-terminal. As demais partes da estrutura contêm regiões de *loops* entre as estruturas principais contendo, em casos esporádicos, algumas folhas- β antiparalelas à folha- β principal.

De posse destes resultados, um representante de cada espécie pertencente a cada uma das subfamílias de NAC identificadas na análise fenética foi selecionado para a modelagem tridimensional das suas respectivas proteínas. Inicialmente foi realizada uma busca por modelos baseada na homologia da sequência de entrada com os modelos disponíveis em bancos de dados como o Swiss-Prot e PDB. Foram selecionadas 4 sequências de soja: Glyma02g12220.4 (GmNAC-a), Glyma04g42800.1 (GmNAC-b), Glyma08g18470.1 (GmNAC-d) e Glyma20g33390.1 (GmNAC-e) para soja e 5 sequências TR59572|c0_g1_i4 (VuNAC-a), TR65034|c8_g3_i2 (VuNAC-b), TR55676|c1_g3_i7 (VuNAC-c), TR121925|c0_g1_i1 (VuNAC-d) e TR73892|c0_g1_i1 (VuNAC-e) para feijão-caupi.

Os resultados da modelagem e confirmação dos modelos revelaram que todas as proteínas modeladas tridimensionalmente apresentaram mais de 92.0 % de resíduos de aminoácidos localizados em regiões estruturalmente ideais, número este dentro do esperado segundo Lovell *et al.* (2002). Os gráficos obtidos (Figuras 10 e 11) demonstram que tais modelos são sustentados matematicamente. Os dímeros modelados, com exceção de GmNAC-a e VuNAC-c (monômeros) estão de acordo com os modelos proteicos demonstrados por outros autores (ERNST *et al.*, 2004; HUSSEY *et al.*, 2015), onde as regiões de ligação com o DNA (*DNA binding domain*) estão representadas pelas folhas- β antiparalelas (destacadas em azul) situadas na região intermediária dos monômeros que compõem os dímeros modelados. Já os monômeros GmNAC-a e VuNAC-c, têm as suas respectivas regiões de ligação ao DNA situadas na região central evidenciadas pelas folhas- β destacadas também em azul (DUVAL *et al.*, 2002; ERNST *et al.*, 2004).

3.3.6 Análise de expressão *in silico* baseada em dados de RNA-seq

Os dados de FPKM obtidos a partir dos bancos de dados de RNA-Seq (estresse radicular) da soja (GENOSOJA) e do feijão-caupi (NordEST) foram clusterizados hierarquicamente com o auxílio do programa Cluster 3.0 e utilizados para gerar o perfil de expressão *in silico* (*heatmap*) das proteínas NAC anotadas (Figura 12). Cinco sequências de soja (Glyma20g31210.1, Glyma04g42800.1, Glyma01g06150.1, Glyma02g12220.4 e Glyma12g35000.1) não foram consideravelmente expressas em nenhuma das bibliotecas, enquanto que cinco sequências Glyma10g36360.1, Glyma06g11970.1, Glyma20g33390.1, Glyma14g24220.1 e Glyma02g26480 foram superexpressas em todas as bibliotecas. Dentre as bibliotecas analisadas em soja, as que menos apresentaram transcritos expressos foram as bibliotecas de nódulos e vagem verde.

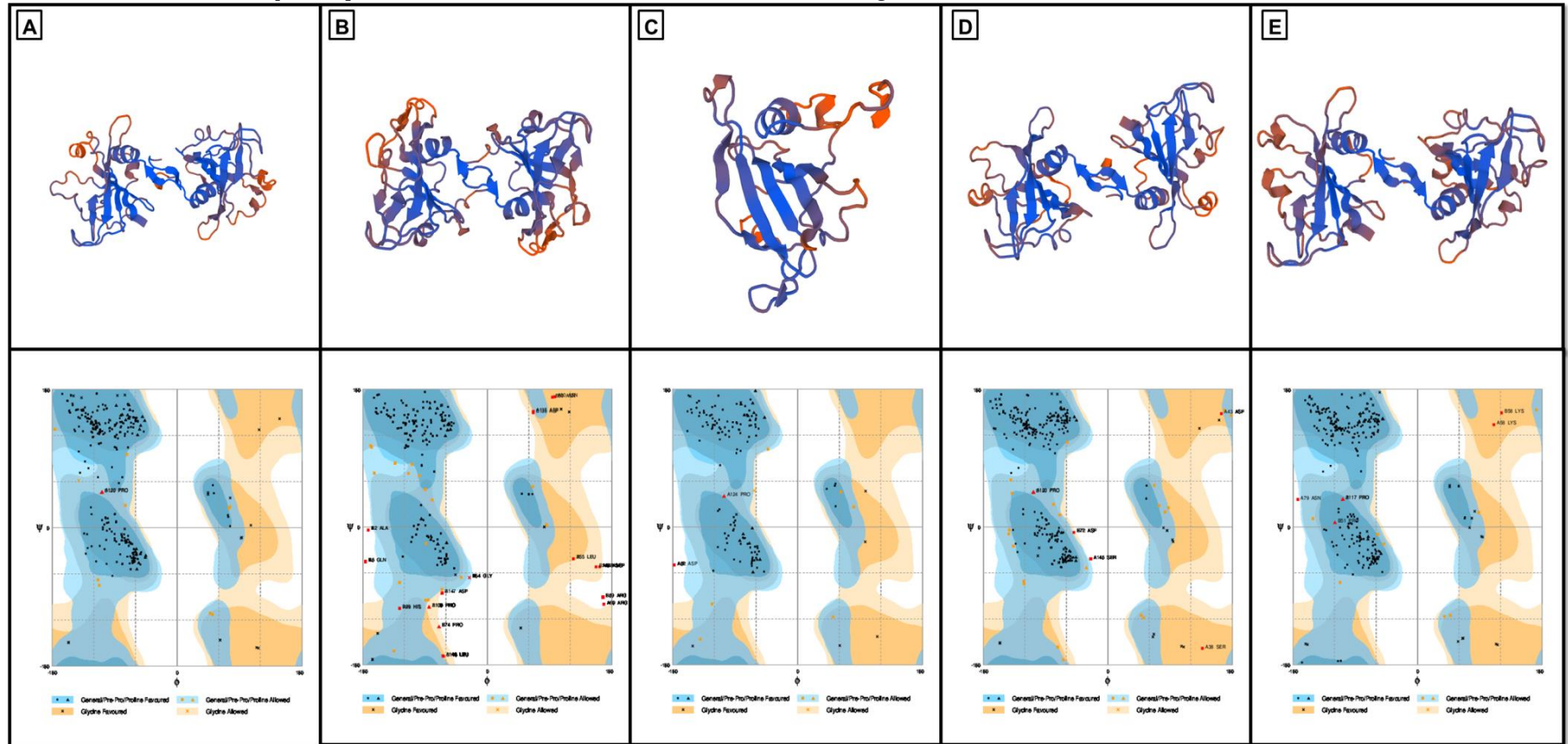
No feijão-caupi é evidente a diferença nos níveis de expressão entre os tempos e as bibliotecas, a exemplo do que se observa em TR59572|c0_g1_i1, TR21378|c0_g1_i1 e TR39042|c0_g2_i1, em que os níveis de expressão são mais evidentes nos estágios mais tardios (150 minutos). Já os transcritos TR121837|c0_g1_i1, TR61551|c1_g2_i1, TR43298|c0_g1_i1, TR43298|c0_g1_i3, TR43298|c0_g2_i1 e TR59572|c0_g1_i5 são superexpressos, praticamente em todas as bibliotecas, especialmente na cultivar tolerante Pingo-de-Ouro.

É sabido que os membros da família NAC são multifuncionais e desempenham papéis importantes em diversos aspectos do desenvolvimento vegetal, que vão desde o desenvolvimento embrionário até a senescência. Além dos aspectos relacionados ao desenvolvimento e crescimento vegetal, estes fatores de transcrição e os seus produtos também são de extrema importância para as vias de resposta imunológica e resistência a estresses bióticos e abióticos (OLSEN *et al.*, 2005; LIU *et al.*, 2014). Por se tratar de um gene tão versátil, espera-se que sua expressão seja observada em diversos tecidos e estágios de desenvolvimento diferentes, bem como sob diferentes condições de estresses bióticos e abióticos. Conforme é possível observar nos resultados obtidos neste estudo, especialmente em soja, algumas das sequências analisadas apresentaram níveis de expressão preferencial nas bibliotecas de estresse biótico, como por exemplo, os transcritos Glyma12g00540.1, Glyma08g17140.1, Glyma09g36829.1, Glyma06g38410.1 e Glyma12g22880.1, que foram superexpressos na maioria das bibliotecas de estresse biótico, enquanto outras apresentaram níveis de expressão preferencial nas bibliotecas de estresse abiótico, como os transcritos Glyma07g05660.1, Glyma16g02200.1, Glyma12g22790.1 e Glyma06g38449.1.

Os membros da subfamília exclusiva do feijão-caupi (NAC-c) e a maioria dos membros das subfamílias mistas NAC-b e NAC-e não apresentaram um padrão preferencial de expressão, tendo apresentado elevados níveis de expressão em todas as bibliotecas, possivelmente por se tratar de genes mais generalistas, contudo, alguns membros das subfamílias NAC-b e NAC-d de soja e NAC-b e NAC-e de feijão-caupi, apresentaram expressão preferencial em raiz.

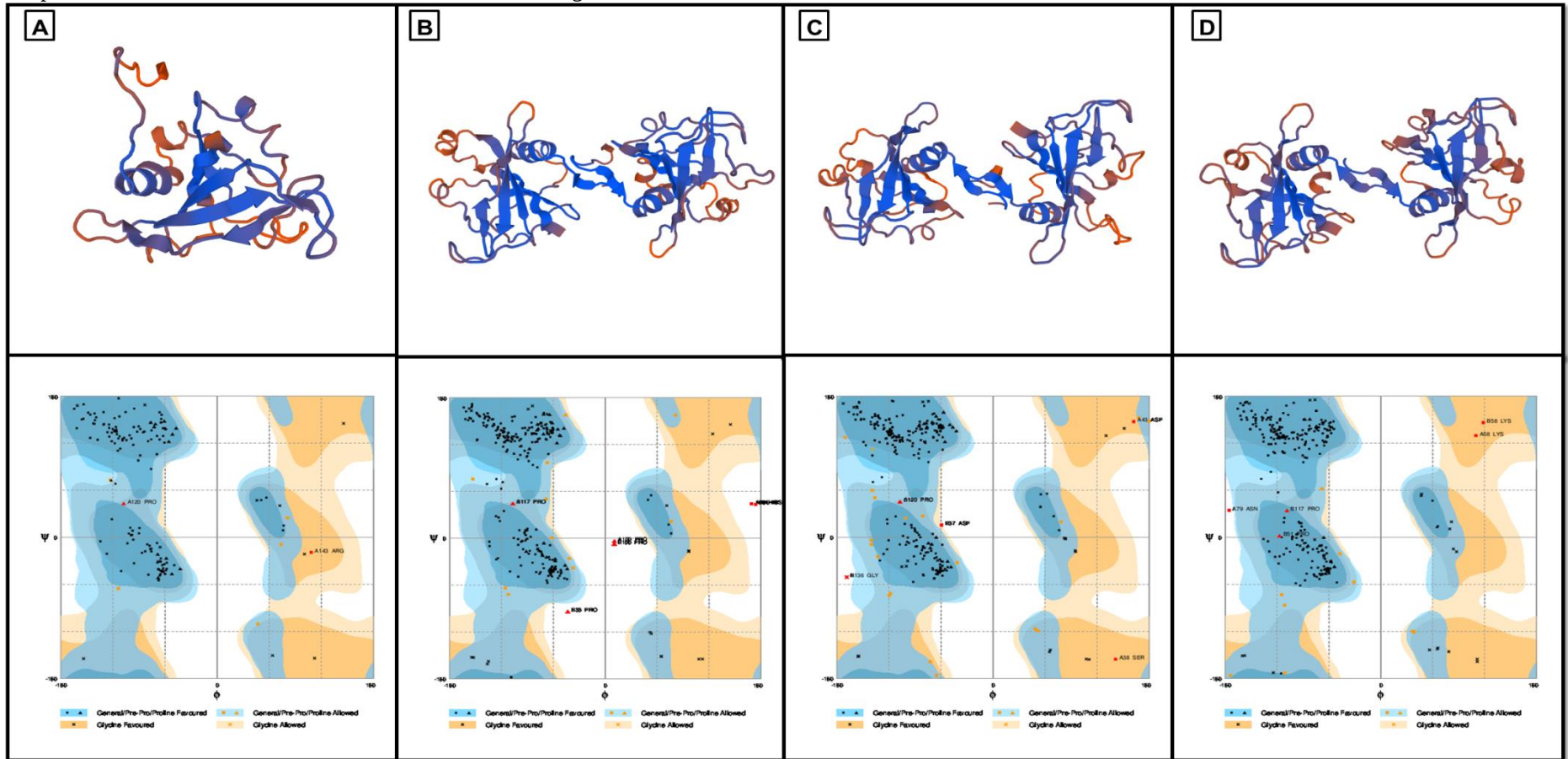
De posse dos dados gerados ao longo das análises, especialmente os dados de *fold-change* das bibliotecas de RNA-Seq, foram selecionadas 10 sequências de soja e 08 de feijão-caupi para servirem como modelo para o desenho de primers (Tabela 6).

Figura 10 — Modelagem tridimensional das proteínas VuNAC-a (A), VuNAC-b (B), VuNAC-c (C), VuNAC-d (D) e VuNAC-e (E) gerado a partir do programa SWISS-MODEL com os seus respectivos gráficos Ramachandran (abaixo). Os gráficos de Ramachandran demonstram, respectivamente, que 99.3%, 92.4%, 98.7%, 98.0% e 98.0% dos resíduos de aminoácidos que compõem as estruturas modeladas encontram-se situados em regiões ideais



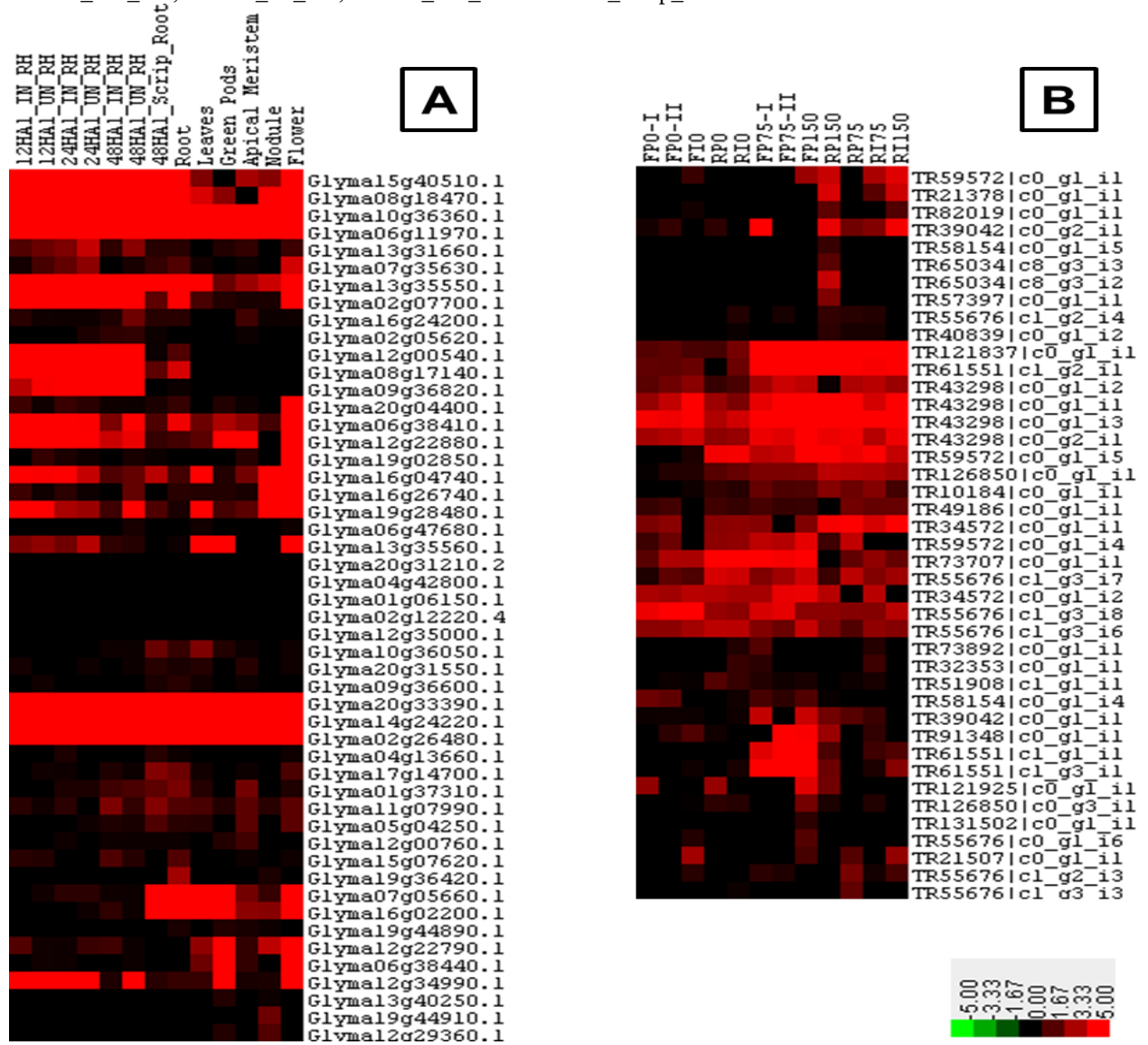
Fonte: o autor (2017).

Figura 11 — Modelagem tridimensional das proteínas GmNAC-a (A), GmNAC-b (B), GmNAC-d (C) e GmNAC-e (D) gerado a partir do programa SWISS-MODEL com seus respectivos gráficos Ramachandran (abaixo). Os gráficos de Ramachandran demonstram, respectivamente, que 98.3%, 97.7%, 98.0% e 98.3% dos resíduos de aminoácidos que compõem as estruturas modeladas encontram-se situados em regiões ideais



Fonte: o autor (2017).

Figura 12 — Perfil de expressão *in silico* (heatmap) para as sequências de soja (A) e feijão-caupi (B) obtidos após clusterização hierárquica realizada pelo programa Cluster 3.0. O gradiente de cores (abaixo à direita) indica o nível de expressão das sequências nas bibliotecas, sendo: ausência de expressão (preto), indução (vermelho) e repressão (verde). Abreviações: Root (raiz), Leaves (folhas), Green Pods (vagens verdes), Apical Meristem (meristema apical), Nodule (nódulo), Flower (flor), FP0-I (Folha Pingo de Ouro 0h – Controle), FP0-II (Folha Pingo de Ouro 0h – Tratamento), F10 (Folha Santo Inácio 0h), RP0 (Raiz Pingo de Ouro 0h), RI0 (Raiz Santo Inácio 0h), FP75-I (Folha Pingo de Ouro 75h – Controle), FP75-II (Folha Pingo de Ouro 75h – Tratamento), FP150 (Folha Pingo de Ouro 150h), RP150 (Raiz Pingo de Ouro 150h), RP75 (Raiz Pingo de Ouro 75h), RI75 (Raiz Santo Inácio 75h), RI150 (Raiz Santo Inácio 150h). As bibliotecas 12HA1_IN_RH, 12HA1_UN_RH, 24HA1_IN_RH, 48HA1_IN_RH, 48HA1_UN_RH e 48HA1_Scrip_Root são bibliotecas de estresse biótico



Fonte: o autor (2017).

Tabela 6 — *Primers* desenhados para soja (Gm_NAC) e feijão-caupi (Vu_NAC) a partir das sequências obtidas nos bancos de dados de RNA-Seq de ambas as espécies, incluindo nome do primer, sequência de entrada, par do primer, posição de início, tamanho do Amplicon, temperatura de anelamento, conteúdo relativo de GC e as sequências

Primer Name	Query	Primer	Start	Amplicon (bp)	Tm (°C)	GC (%)	Sequence
Gm_NAC_01	Glyma10g36360.1	Glyma10g36360.1_F	1258	171	60	50	AGAAACACCCCAAGTGCATC
		Glyma10g36360.1_R	1428		60	55	CTCAACAAAGGCCCAGAGAG
Gm_NAC_02	Glyma06g11970.1	Glyma06g11970.1_F	158	195	59.9	50	CCGGATTGGCTACTTATGGA
		Glyma06g11970.1_R	352		60	50	TAGGAGCTTTCCTGCGTAA
Gm_NAC_03	Glyma06g38410.1	Glyma06g38410.1_F	422	189	60	45	ATTCTTCCCGCAAACACAAC
		Glyma06g38410.1_R	610		59.9	45	CATTTATCTCCGGCAACGAT
Gm_NAC_04	Glyma13g35550.1	Glyma13g35550.1_F	687	185	59.9	50	TGTTTCAGAACTTGGGTGCAG
		Glyma13g35550.1_R	871		59.9	50	TTAACGGAACCGATGAGTCC
Gm_NAC_05	Glyma01g22510.1	Glyma01g22510.1_F	426	231	59.9	45	AGGGCAAGATTTTGGTGATG
		Glyma01g22510.1_R	656		60.1	50	TGCTGCTGCTGGTAAAACAC
Gm_NAC_06	Glyma02g11140.1	Glyma02g11140.1_F	637	168	60.1	50	GTGTTTTACCAGCAGCAGCA
		Glyma02g11140.1_R	804		60	55	ACAAGGAGGCTCAGGAGTCA
Gm_NAC_07	Glyma19g28480.1	Glyma19g28480.1_F	317	219	60.1	50	AGCAACCATCTCCAATCGAC
		Glyma19g28480.1_R	535		60	50	AGTTCCATTGTGGGTCTCTCG
Gm_NAC_08	Glyma16g04740.1	Glyma16g04740.1_F	658	185	59.9	45	TTGTTGGATGCCATGGACTA
		Glyma16g04740.1_R	842		60	45	GGGTTCAATTGGGGATTCTT
Gm_NAC_09	Glyma12g34990.1	Glyma12g34990.1_F	509	167	60	50	TGCTGTGCCGAATCTACAAG
		Glyma12g34990.1_R	675		59.9	55	GCTGTTGTAGGCTGTGGTCA

Gm_NAC_10	Glyma14g24220.1	Glyma14g24220.1_F	434	193	59.8	55	GCTTAAGGCTGGATGACTGG
		Glyma14g24220.1_R	626		59.8	55	GCCTCGAAGTACAGGCAATC
Vu_NAC_01	TR121837 c0_g1_i1	TR121837 c0_g1_i1_F	1359	220	60.1	40	GCCGAAAATGAAGACCAAAA
		TR121837 c0_g1_i1_R	1578		60	55	ACGCAGGGGTATCAGAACAC
Vu_NAC_02	TR61551 c1_g1_i1	TR61551 c1_g1_i1_F	208	222	60.1	45	TGGGCACCATTTTTCTCTTC
		TR61551 c1_g1_i1_R	429		60	50	ACTTTTCTCCCTTCGGTGGT
Vu_NAC_03	TR73707 c0_g1_i1	TR73707 c0_g1_i1_F	713	155	60.1	45	CCCTTCAATGCGAAACTCAT
		TR73707 c0_g1_i1_R	867		59.9	60	CTGCCTCAGGGTACTGGAAG
Vu_NAC_04	TR126850 c0_g3_i1	TR126850 c0_g3_i1_F	232	150	60.2	45	TTTATGGGCGAAGCTTATCG
		TR126850 c0_g3_i1_R	381		59.9	55	TGTAGCAGACTCGGGGAAGT
Vu_NAC_05	TR51908 c1_g1_i1	TR51908 c1_g1_i1_F	527	222	60.1	45	CTCCGAATGGCATCAAAACT
		TR51908 c1_g1_i1_R	748		60	50	AGGAAGGGAAAGAGGTGGAA
Vu_NAC_06	TR59572 c0_g1_i1	TR59572 c0_g1_i1_F	1	215	60	45	CTTCACCATTTTTGGCACCT
		TR59572 c0_g1_i1_R	215		59.9	50	GAAACCTGAAACCAGGTGGA
Vu_NAC_07	TR21507 c0_g1_i1	TR21507 c0_g1_i1_F	1573	169	60	55	CTCTCTGGGCCTTTGTTGAG
		TR21507 c0_g1_i1_R	1741		60.1	45	TTTCTGCCTGCTTTCTTCGT
Vu_NAC_08	TR55676 c1_g3_i7	TR55676 c1_g3_i7_F	198	226	60	50	ATAACCCCTTTTCCCTCCT
		TR55676 c1_g3_i7_R	423		60.1	45	GCTCCCCATTCTTCACAAAA

Fonte: o autor (2017).

Os pares de *primers* foram desenhados especificamente para suas respectivas espécies, tendo como objetivo a posterior validação da expressão nos transcriptomas da soja e do feijão-caupi via RT-qPCR.

3.4 CONCLUSÕES

Este estudo permitiu a identificação e caracterização de 42 membros da família NAC em feijão-caupi e 50 em soja, analisando estes resultados comparativamente com os dados de soja.

Os métodos empregados foram eficientes na descrição estrutural e funcional destes genes, bem como na identificação das relações fenéticas, motivos conservados e suas respectivas composições, assim como os padrões de expressão e estrutura proteica.

Os dados fornecem informações úteis para estudos mais aprofundados relacionados às funções específicas destes fatores de transcrição em soja e feijão-caupi, podendo ser usados no desenvolvimento de marcadores moleculares para o melhoramento vegetal visando uma maior tolerância destas espécies e seus cultivares.

REFERÊNCIAS

- AIDA, M.; ISHIDA, T.; FUKAKI, H.; FUJISAWA, H.; TASAKA, M. Genes involved in organ separation in *Arabidopsis*: an analysis of the *cup-shaped cotyledon* mutant. **Plant Cell**, v. 9, p. 841-857, 1997.
- ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The SWISS-MODEL Workspace: A web-based environment for protein structure modelling. **Bioinformatics**, v. 22, p. 195-201, 2006.
- BAILEY, T.L.; BODÉN, M.; BUSKE, F.A.; FRITH, M.; GRANT, C.E.; CLEMENTI, L.; REN, J.; LI, W.W.; NOBLE, W.S. MEME SUITE: tools for motif discovery and searching. **Nucleic Acids Research**, v. 37, p. 202-208, 2009.
- BAILEY, T.L.; ELKAN, C. **Fitting a mixture model by expectation maximization to discover motifs in biopolymers**. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. 1994. p. 28-36
- BASINI, M.; BIENERT, S.; WATERHOUSE, A.; ARNOLD, K.; STUDER, G.; SCHMIDT, T.; KIEFER, F.; CASSARINO, T.G.; BERTONI, M.; BORDOLI, L.; SCHWEDE, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, v. 42, p. 252-258, 2014.

BECK, E.H.; FETTIG, S.; KNAKE, C.; HARTING, K.; BHATTARAI, T. Specific and unspecific responses of plants to cold and drought stress. **Journal of Biosciences**, v. 32, n. 3, p.501-510, 2007.

BRAY, E. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. **Journal of Experimental Botany**, v. 55, n. 407, 2004.

DO VALE, N.M.; BARILI, L.D.; ROZZETO, D.S.; STINGHIN, J.C.; COIMBRA, J.L.M.; GUIDOLIN, A.F.; KÖOP, M.M. Avaliação para tolerância ao estresse hídrico em feijão. **Biotemas**, v. 25, n. 3, p.135-144, 2012.

DROZDETSKIY, A.; COLE, C.; PROCTER, J.; BARTON, G.J. JPred4: a protein secondary structure prediction server. **Nucleic Acids Research**, v. 43, p. 389-394, 2015.

DUVAL, M.; HSIEH, T.F.; KIM, S.Y.; THOMAS, T.L. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. **Plant Molecular Biology**, v. 50, p. 237-248, 2002.

EISEN, M.B.; SPELLMAN, P.T.; BROWN, P.O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**, USA, v. 95, p. 14863-14868, 1998.

ERNST, H.; OLSEN, A.; LARSEN, S.; LEGGIO, L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. **EMBO reports**, v. 5, p. 297-303, 2004.

FANG, Y.; YOU, J.; XIE, K.; XIE, W.; XIONG, L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. **Molecular Genetics and Genomics**, v. 280, p. 547-563, 2008.

GUDEX, N.; PEITSCH, M.C.; SCHWEDE, T. Automated comparative protein structure modelling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. **Electrophoresis**, v. 30, n. 1, p. 162-173, 2009.

HU, H.; DAI, M.; YAO, J.; XIAO, B.; LI, X.; ZHANG, Q. Overexpressing a NAM, ATAF and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. **Proceedings of the National Academy of Sciences**, USA, v. 103, p. 12987-12992, 2006.

HU, R.; QI, G.; KONG, Y.; KONG, D.; GAO, Q.; ZHOU, G. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. **BMC Plant Biology**, v. 10, p. 145, 2010.

JENSEN, M.; KJAERGAARD, T.; NIELSEN, M.; GALBERG, P.; PETERSEN, K.; O'SHEA, C.; SKRIVER, K. The *Arabidopsis thaliana* NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. **Biochemical Journal**, v. 426, p. 183–196, 2010.

JIN, H.; XU, G.; MENG, Q.; HUANG, F.; YU, D. GmNAC5, a NAC transcription factor, is a transient response regulator induced by abiotic stress in soybean. **Scientific World Journal**, v. 2013, p.768-972, 2013.

KASUGA, M.; MIURA, S.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. A combination of the *Arabidopsis* DREB1A gene and stress-inducible *rd29a* promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. **Plant and Cell Physiology**, v. 45, n. 3, p.346-350, 2004.

KIEFER, F.; ARNOLD, K.; KÜNZLI, M.; BORDOLI, L.; SCHWEDE, T. The SWISS-MODEL Repository and associated resources. **Nucleic Acids Research**, v. 37, p.387-392, 2009.

KIKUCHI, K.; UEGUSHI-TANAKA, M.; YOSHIDA, K.T.; NAGATO, Y.; MATSUSOKA, M.; HIRANO, H.Y. Molecular analysis of the NAC gene family in rice. **Molecular Genomics and Genetics**, v. 262, p. 1047-1951, 2000.

KIM, S.; KIM, S.; PARK, C. A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUST in *Arabidopsis*. **Planta**, v. 226, p. 647-654, 2007.

LIU, T.; SONG, X.; DUAN, W.; HUANG, Z.; LIU, G.; LI, Y. Genome-wide analysis and expression patterns of NAC transcription factor family under different developmental stages and abiotic stresses in Chinese cabbage. **Plant Molecular Biology Reporter**, v. 32, p. 1041-1056, 2014.

LIU, T.; ZHU, S.; TANG, Q.; TANG, S. Identification of 32 full-length NAC transcription factors in ramie (*Boehmeria nivea* L. Gaud) and characterization of the expression pattern of these genes. **Molecular Genetics and Genomics**, v. 289, p. 675-684, 2014.

MANTRI, N.; PATADE, V.; PENNA, S.; FORD, R.; PANG, E. Abiotic Stress Response in Plants: Present and Future. In: AHMAD, P.; PRASAD, M.N.V. **Abiotic Stress Response in Plants: Metabolism, Productivity and Sustainability**. Springer Science+Business Media. LLC, 2012. p.18-36

MARCHLER-BAUER, A.; DERBYSHIRE, M.K.; GONZALES, N.R.; LU, S.; CHITSAZ, F.; GEER, L.Y.; GEER, R.C.; HE, J.; GWARDZ, M.; HURWITZ, D.I.; LANCZYCKI, C.J.; LU, F.; MARCHLER, G.H.; SONG, J.S.; THANKI, N.; WANG, Z.; YAMASHITA, R.A.; ZHANG, D.; ZHENG, C.; BRYANT, S.H. CDD: NCBI's conserved domain database. **Nucleic Acids Research**, v. 43, 2015.

NAKASHIMA, K.; TAKASAKI, H.; UNYAMIZOI, J.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. NAC transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta**, v. 1819, p. 97-103, 2012.

NURUZZAMAN, M.; MANIMEKALAI, R.; SHARONI, A.; SATOH, K.; KONDOH, H.; OOKA, H. Genome-wide analysis of NAC transcription factor family in rice. **Gene**, v. 465, p. 30-44, 2010.

OLSEN, A.N.; ERNST, H.A.; LEGGIO, L.L.; SKRIVER, K. NAC transcription factors: structurally distinct, functionally diverse. **Trends in Plant Science**, v. 10, p. 79-87, 2005.

OOKA, H.; SATOH, K.; DOI, K.; NAGATA, T.; OTOMO, Y.; MURAKAMI, K.; MATSUBARA, K.; OSATO, N.; KAWAI, J.; CARNINCI, P. Comprehensive analysis of

NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. **DNA Research**, v. 10, p. 239–247, 2003.

PELLEGRINESCHI, A.; REYNOLDS, M.; PACHECO, M.; BRITO, R.M.; ALMERAYA, R.; YAMAGUCHI-SHINOZAKI, K.; HOISINGTON, D. Stress-induced expression in wheat of the *Arabidopsis thaliana* DREB1A gene delays water stress symptoms under greenhouse conditions. **Genome**, v. 47, p. 493-500, 2004.

PINHEIRO, G.L.; MARQUES, C.S.; COSTA, M.D.; REIS, P.A.; CARVALHO, C.M.; FIETTO, L.G.; FONTES, E.P. Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. **Gene**, v. 444, n. 1-2, p.10-23, 2009.

RAMALHO, M.A.P.; SILVA, G.S.; DIAS, L.A.S. Genetic plant improvement and climate changes. **Crop Breeding and Applied Biotechnology**, v. 9, n. 2, p. 189-195, 2009.

SAITOU, N.; NEI, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, p. 406-425, 1987.

SALDANHA, A.J. Java Treeview – extensible visualization of microarray data. **Bioinformatics**, v. 20, n. 17, p. 3246-3248, 2004.

SCHWARZ, R.; DAYHOFF, M. Matrices for detecting distant relationships. In: DAYHOFF, M. **Atlas of protein sequences**.1979. p.353-358.

SIEVERS, F.; WILM, A.; DINEEN, D.; GIBSON, T.J.; KARPLUS, K.; LI, W.; LOPEZ, R.; MCWILLIAM, H.; REMMERT, M.; SÖDING, J.; THOMPSON, J.D.; HIGGINS, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, p. 539, 2011.

SOUER, E.; VAN HOUWELINGEN, A.; KLOOS, D.; MOL, J.; KOES, R. The *No Apical Meristem* gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. **Cell**, v. 85, p. 159-170, 1996.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. **Molecular Biology and Evolution**, v. 30, p. 2725-2729, 2013.

TAOKA, K.; YANAGIMOTO, Y.; DAIMON Y.; HIBARA, K.; AIDA, M.; TASAKA, M. The NAC domain mediates functional specificity of CUP-SHAPED COTYLEDON proteins. **Plant Journal**, v. 40, p. 462-473, 2004.

TRAN, L.; NAKASHIMA, K.; SAKUMA, Y.; SIMPSON, S.; FUJITA, Y.; MARUYAMA, K. Isolation and functional analysis of *Arabidopsis* stress inducible NAC transcription factors that bind to a drought responsive cis-element in the early responsive dehydration stress 1 promoter. **Plant Cell**, v. 18, p. 2481-2498, 2004.

WU, J.; WANG, L.; WANG, S. Comprehensive analysis and discovery of drought-related NAC transcription factors in common bean. **BMC Plant Biology**, v. 16, p. 193-205, 2016.

XU, B.; OHTANI, M.; YAMAGUSHI, M.; TOYOOKA, K.; WAKAZAKI, M.; SATO, M. Contribution of NAC transcription factors to plant adaptation to land. **Science**, v. 343, p. 1505-1508, 2014.

YAMAGUCHI-SHINOZAKI, K.; KASUGA, M.; LIU, Q.; NAKASHIMA, K.; SAKUMA, Y.; ABE, H.; SHIWARI, Z.K.; SEKI, M.; SHINOZAKI, L. Biological mechanisms of drought stress response. **JIRCAS Working Report**, v. 23, p. 1-8, 2002.

ZHENG, X.; TANG, S.; ZHU, S.; DAI, Q.; LIU, T. Identification of a NAC Transcription Factor Family by Deep Transcriptome Sequencing in Onion (*Allium cepa* L.). **Plos One**, v. 11, n. 6, 2016.

4 CONCLUSÕES

- Os genes da família NAC estão presentes no transcriptoma da soja e do feijão-caupi, tendo sido identificados e caracterizados 42 e 50 membros da família NAC nos transcriptomas de feijão-caupi e de soja, respectivamente;
- Foram identificados 20 motivos conservados nas sequências de soja e feijão-caupi, cujos padrões de distribuição e composição de aminoácidos se mantiveram conservados em ambas as espécies;
- A maioria dos genes analisados, conforme esperado, encontra-se situada no núcleo. Contudo, alguns genes não apresentaram os padrões habituais de sinalização nuclear, conforme descrito anteriormente por alguns autores;
- Foram identificadas sequências correspondentes às cinco subfamílias NAC em feijão-caupi, enquanto na soja apenas quatro apresentaram representantes, não havendo representantes da subfamília “NAC-c”;
- Ao analisar a composição das sequências específicas de cada subfamília, foi possível observar um maior nível de conservação e similaridade na região 5’ e uma relativa variação na região 3’;
- As análises de identificação de estrutura secundária demonstraram que os resultados obtidos estão de acordo com o descrito na literatura e em associação com os resultados da modelagem tridimensional e de Ramachandran, foi possível gerar modelos aceitáveis de algumas das proteínas NAC analisadas, sendo possível, inclusive, identificar os sítios de ligação ao DNA nos modelos gerados;
- A análise da expressão dos FTs NAC revelou alguns candidatos superexpressos em todos tecidos e situações (inclusive controles não estressados), enquanto outros apresentaram baixa expressão ou especificidade de tempo e tecido. Os membros das subfamílias NAC-b e NAC-d (em soja) e NAC-b e NAC-e (em feijão-caupi), apresentaram expressão preferencial em raiz.

REFERÊNCIAS

- AGUIAR, R. S.; MODA-CIRINO, V.; FARIA, R. T.; VIDAL, L. H. I. Avaliação de linhagens promissoras de feijoeiro (*Phaseolus vulgaris* L.) tolerantes ao déficit hídrico. **Semina: Ciências Agrárias**, Londrina, v. 29, n. 1, p. 1-14, 2008.
- AIDA, M.; ISHIDA, T.; FUKAKI, H.; FUJISAWA, H.; TASAKA, M. Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. **Plant Cell**, v. 9, p. 841-857, 1997.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic Local Alignment Search Tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.
- AMORIM, L. D. M.; SOUSA, L. O. F.; OLIVEIRA, F. F. M.; CAMACHO, R. G. V.; MELO, J. I. M. Fabaceae na floresta nacional (FLONA) de Assú, Semiárido potiguar, Nordeste do Brasil. **Rodriguésia**, v. 67, n. 1, p. 105-123, 2016.
- AMORIM, L. L. B.; FONSÊCA-DOS-SANTOS, R.; BEZERRA-NETO, J. P.; GUIDA-SANTOS, M.; CROVELLA, S.; BENKO-ISEPPON, A. M. Transcription factors involved in plant resistance to pathogens. **Current Protein & Peptide Science**, v. 18, n. 4, p. 335-351, 2017.
- ANDRADE JUNIOR, A. S.; RODRIGUES, B. H. N.; FRIZZONE, J. A.; CARDOSO, M. J.; BASTOS, E. A.; MELO, F. B. Níveis de irrigação na cultura do feijão-caupi. **Revista brasileira de engenharia agrícola e ambiental**, v. 6, n. 1, p. 17-20, 2002.
- ANDRADE, R. P.; KARIA, C. T. **O uso de *Stylosanthes* em pastagens no Brasil**. In: SIMPÓSIO DE FORRAGICULTURA E PASTAGEM, 1., 2000, Lavras. Temas em evidências. Lavras: UFLA, p. 173-309, 2000.
- APEL, K.; HIRT, H. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. **Annual Review of Plant Biology**, v. 55, p. 373-399, 2004.
- APG III. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. **Botanical Journal of the Linnean Society**, v. 161, p. 105-121, 2009.
- ARAÚJO, N. D.; FARIAS, R. P.; PEREIRA, P. B.; FIGUEIRÊDO, F. M.; MORAIS, A. M. B.; SALDANHA, L. C.; GABRIEL, J. E. A era da bioinformática: Seu potencial e suas implicações para as ciências da saúde. **Estudos de Biologia**, v. 30, p. 143-148, 2008.
- ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The SWISS-MODEL Workspace: A web-based environment for protein structure modelling. **Bioinformatics**, v. 22, p. 195-201, 2006.
- ASAMAA, K.; SOBER, A.; HARTUNG, W.; NINEMETS, U. Rate of stomatal opening, shoot hydraulic conductance and photosynthetic characteristics in relation to leaf abscisic acid concentration in six temperate deciduous trees. **Tree Physiology**, v. 22, p. 267-276, 2002.

BAILEY, T. L.; BODÉN, M.; BUSKE, F. A.; FRITH, M.; GRANT, C. E.; CLEMENTI, L.; REN, J.; LI, W. W.; NOBLE, W. S. MEME SUITE: tools for motif discovery and searching. **Nucleic Acids Research**, v. 37, p. 202-208, 2009.

BAILEY, T. L.; ELKAN, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. **Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology**, p. 28-36, 1994.

BARBA, M.; CZOSNEK, H.; HADIDI, A. Historical perspective, development and applications of Next-Generation Sequencing in plant virology. **Viruses**, v. 6, p. 106-136, 2014.

BARNABÁS, B.; JAGER, K.; FEHÉR, A. The effect of drought and heat stress on reproductive processes in cereals. **Plant, Cell and Environment**, v. 31, p. 11-38, 2008.

BARTELS, D.; SUNKAR, R. Drought and salt tolerance in plants. **Critical Review in Plant Sciences**, v. 24, p. 23-58, 2005.

BASINI, M.; BIENERT, S.; WATERHOUSE, A.; ARNOLD, K.; STUDER, G.; SCHMIDT, T.; KIEFER, F.; CASSARINO, T. G.; BERTONI, M.; BORDOLI, L.; SCHWEDE, T. Swiss-model: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, v. 42, p. 252-258, 2014.

BAUDOIN, J.P.; MARÉCHAL, R. Genetic Diversity in Vigna. In: SINGH, S.R.; RACHIE, K.O. (Ed.). **Cowpea Research, Production and Utilization**. Chichester, U.K.: John Wiley & Sons, p. 11-21, 1985.

BAXEVANIS, A. D.; OUELLETTE, B. F. F. **Bioinformatics: a practical guide to the analysis of genes and proteins**. 3 ed. Hoboken, N.J.: John Wiley & Sons, 2005.

BECK, E. H.; FETTIG, S.; KNAKE, C.; HARTING, K.; BHATTARAI, T. Specific and unspecific responses of plants to cold and drought stress. **Journal of Biosciences**, v. 32, n. 3, p. 501-510, 2007.

BEEBE, S. E.; IDUPULATPATI, R. A. O.; MATTHEW, W. B.; BUTARE, L. Selection for drought resistance in common bean also improves yield in phosphorus limited and favorable environments. **Crop Science**, Madison, v. 48, n. 2, p. 582-592, 2008.

BENKO-ISEPPON, A. M.; SOARES-CAVALCANTI, N. M.; BELARMINO, L. C.; BEZERRA-NETO, J. P.; AMORIM, L. L. B.; FERREIRA-NETO, J. R. C.; PANDOLFI, V.; AZEVEDO, H. M. A.; SILVA, R. L. O.; SANTOS, M. G.; ALVES, M. V.; KIDO, E. A. Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas. **Revista Brasileira de Geografia Física**, v. 6, p. 1112-1134, 2011.

BFG (BRAZIL FLORA GROUP). Growing knowledge: an overview of Seed Plant diversity in Brazil. **Rodriguésia**, v. 66, p. 1085-1113, 2015.

BOECKMANN, B.; BAIROCH, A.; APWEITER, R.; BLATTER, M. C.; ESTREICHER, A.; GASTEIGER, E.; MARTIN, M. J.; MICHOD, K.; O'DONNOVAN, C.; PHAN, I.;

PILBOUT, S.; SCHNEIDER, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Research**, v. 31, n. 1, p. 365-370, 2003.

BOHNERT, H. J.; NELSON, D. E.; JENSEN, R. G. Adaptations to environmental stresses. **Plant Cell**, v. 7, p. 1099-1111, 1995.

BONATO, E. R. **Estresse em Soja**. Passo Fundo: EMBRAPA Trigo, p. 254, 2000.

BRAY, E. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. **Journal of Experimental Botany**, v. 55, n. 407, 2004.

CHAVES, M. M.; MAROCO, J. P.; PEREIRA, J. S. Understanding plant responses to drought – from genes to the whole plant. **Functional Plant Biology**, v. 30, p. 239-264, 2003.

CHEN, L.; SONG, Y.; LI, S.; ZHANG, L.; ZOU, C.; YU, D. The role of WRKY transcription factors in plant abiotic stresses. **Biochimica et Biophysica Acta**, v. 1819, p. 120-128, 2012.

CHINNUSAMY, V.; XIONG, L.; ZHU, J. K. Use of Genetic engineering and molecular biology approaches for crop improvements. In: ASRAF, M.; HARRIS, P.J.C. (Ed). **Abiotic Stresses: plant resistance through breeding and molecular approaches**. New York: Food Products Press, p. 47-95, 2005. cap. 2

COLLINGE, M.; BOLLER, T. Differential induction of two potato genes, Stprx2 and StNAC, in response to infection by *Phytophthora infestans* and to wounding. **Plant Molecular Biology**, v. 46, p. 521-529, 2001.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). Acompanhamento da Safra Brasileira – Décimo Levantamento. **Boletim de grãos**, v. 3, n. 10, 2016.

COMSTOCK, J. P. Hydraulic and chemical signaling in the control of stomatal conductance and transpiration. **Journal of Experimental Botany**, v. 53, p. 195-200, 2002.

CORDEIRO, L. G.; BEZERRA, F. M. L.; SANTOS, J. J. A.; MIRANDA, E. P. Fator de sensibilidade ao déficit hídrico da cultura do feijão caupi (*Vigna unguiculata* (L.) walp.). **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 2, n. 2, p. 153-157, 1998.

CORNIC, G. Drought stress inhibits photosynthesis by decreasing stomatal aperture – not by affecting ATP synthesis. **Trends in Plant Science**, v. 5, p. 187-188, 2000.

COSTA NETO, P. R.; ROSSI, L. F. S. Produção de biocombustível alternativo ao óleo diesel através da transesterificação de óleo de soja usado em fritura. **Química Nova**, v. 23, p. 4, 2000.

DELESSERT, C.; KAZAN, K.; WILSON, I. W.; VAN DER STRAETEN, D.; MANNERS, J.; DENNIS, E. S.; DOLFERUS, R. The transcription factor ATAF2 represses the expression of pathogenesis-related genes in *Arabidopsis*. **The Plant Journal**, v. 43, n. 5, p. 745-757, 2005.

DELLA COLETTA, L. **Estudo da fixação biológica de nitrogênio em leguminosas (família Fabaceae) arbóreas tropicais através do enriquecimento isotópico do ¹⁵N**.

Dissertação (Mestrado em Ciências) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2010.

DIXON, R. A.; HARRISON, M. J. Early events in the activation of plant defense responses. **Annual Review of Phytopathology**, v. 32, p. 479-501, 1994.

DO VALE, N. M.; BARILI, L. D.; ROZZETO, D. S.; STINGHIN, J. C.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; KÖOP, M. M. Avaliação para tolerância ao estresse hídrico em feijão. **Biotemas**, v. 25, n. 3, p. 135-144, 2012.

DROZDETSKIY, A.; COLE, C.; PROCTER, J.; BARTON, G. J. JPred4: a protein secondary structure prediction server. **Nucleic Acids Research**, v. 43, p.389-394, 2015.

DURRANT, W. E.; ROWLAND, O.; PIEDRAS, P.; HAMMOND-KOSACK, K. E.; JONES, J. D. G. cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles. **Plant Cell**, v. 12, p. 963-977, 2000.

DUVAL, M.; HSIEH, T. F.; KIM, S. Y.; THOMAS, T. L. Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily. **Plant Molecular Biology**, v. 50, p. 237-248, 2002.

EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences, USA**, v. 95, p.14863-14868, 1998.

EMBRAPA. **Cultivo do Feijão-caupi em Sistema Agrícola Familiar**. Circular Técnica, 2011.

ERNST, H.; OLSEN, A.; LARSEN, S.; LEGGIO, L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. **EMBO Reports**, v. 5, p. 297-303, 2004.

Eyrich, V. A.; Marti Renom, M. A.; Przybylski, D.; Madhusudhan, M. S.; Fiser, A.; Pazos, F.; Valencia, A.; Sali, A.; Rost, B. **EVA: continuous automatic evaluation of protein structure prediction servers**. *Bioinformatics*, v. 17, p. 1242-1243, 2001.

FANG, Y.; YOU, J.; XIE, K.; XIE, W.; XIONG, L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. **Molecular Genetics and Genomics**, v. 280, p. 547-563, 2008.

FERREIRA, B. M.; COSTA, S. M. N. O gênero *Stylosanthes* Sw. no Brasil. **Epaming**, Belo Horizonte, p. 170, 1979.

FRANCO, A. A.; FARIA, S. M. The contribution of N₂-fixing tree legumes to land reclamation and sustainability in the tropics. **Soil biology and Biochemistry**, v. 29, n. 5/6, p. 897-903, 1997.

FREIRE FILHO, F. R. Feijão-caupi no Brasil: Produção, melhoramento genético, avanços e desafios. **EMBRAPA meio Norte**, p. 84, 2011.

FUJITA, M.; FUJITA, Y.; NOUTOSHI, Y., TAKAHASHI, F.; NARUSAKA, Y.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. **Current Opinion in Plant Biology**, v. 9, p. 436-442, 2006.

GALLOWAY, J. N.; DENTENER, F. J.; CAPONE, D. G.; BOYER, E. W.; HOWARTH, R. W.; SEITZINGER, S. P.; ASNER, G. P.; CLEVELAND, C. C.; GREEN, P. A.; HOLLAND, E. A.; KARL, D. M.; MICHAELS, A. F.; PORTER, J. H.; TOWNSEND, A. R.; VOROSMARTY, C. J. Nitrogen cycles: past, present and future. **Biogeochemistry**, v. 70, p. 153-226, 2004.

GAO, J. P.; CHAO, D. Y.; LIN, H. X. Understanding Abiotic Stress Tolerance Mechanisms: Recent Studies on Stress Response in Rice. **Journal of Integrative Plant Biology**, v. 49, n. 6, p. 742-750, 2007.

GARCIA, J. R.; BUAINAIN, A. M. Evolução recente do agronegócio no cerrado nordestino. **Estudos Sociedade e Agricultura**, v. 23, n. 1, p. 166-195, 2015.

GRANT, E. P.; PICKARD, M. D.; BRISKIN, M. J.; GUTIERREZ-RAMOS, J. C. Gene expression profiles: Creating new perspectives in arthritis research. **Arthritis & Rheumatology**, v. 46, n. 4, p. 874-884, 2002.

GUERX, N.; PEITSCH, M. C.; SCHWEDE, T. Automated comparative protein structure modelling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. **Electrophoresis**, v. 30 (S1): p. 162-173, 2009.

HÄUPL, T.; KRENN, V.; STUHLMÜLLER, B.; RADBRUCH, A.; BURMESTER, G. Perspectives and limitations of gene expression profiling in rheumatology: New molecular strategies. **Arthritis Research & Therapy**, v. 6, n. 4, p. 140-146, 2004.

HU, H.; DAI, M.; YAO, J.; XIAO, B.; LI, X.; ZHANG, Q.; XIONG, L. Overexpressing a NAM, ATAF and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. **Proceedings of the National Academy of Sciences**, USA, v. 103, n. 35, p. 12987-12992, 2006.

HU, R.; QI, G.; KONG, Y.; KONG, D.; GAO, Q.; ZHOU, G. Comprehensive Analysis of NAC Domain Transcription Factor Gene Family in *Populus trichocarpa*. **BMC Plant Biology**, v. 10, p. 145, 2010.

HUH, S. U.; LEE, S. B.; KIM, H. H.; PAEK, K. H. ATAF2, a NAC transcription factor, binds to the promoter and regulates NIT2 gene expression involved in auxin biosynthesis. **Molecules and Cells**, v. 34, p. 305-313, 2012.

HURA, T.; HURA, K.; GRZESIAK, M.; RZEPKA, A. Effect of long-term drought stress on leaf gas exchange and fluorescence parameters in C₃ and C₄ plants. **Acta Physiologiae Plantarum**, v. 29, p. 103-113, 2007.

HUSSEY, S. G.; SAÏDI, M. N.; HEFER, C. A.; MYBURG, A. A. Grima-Pettenati, J. Structural, evolutionary and functional analysis of the NAC domain protein family in *Eucalyptus*. **New Phytologist**, v. 206, p. 1337-1350, 2015.

ISODA, A.; WANG, P. Leaf temperature and transpiration of field-grown cotton and soybean under arid and humid conditions. **Plant Production Science**, v. 5, p. 224-228, 2002.

JENSEN, M. K.; KJARSGAARD, T.; NIELSEN, M. M.; GALBERG, P.; PETERSEN, K.; O'SHEA, C.; SKRIVER, K. The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. **Biochemical Journal**, v. 426, n. 2, p. 183-196, 2010.

JENSEN, M. K.; SKRIVER, K. NAC Transcription Factor Gene Regulatory and Protein-Protein Interaction Networks in Plant Stress Responses and Senescence. **International Union of Biochemistry and Molecular Biology**, v. 66, n. 3, p. 156-166, 2014.

JIN, H.; XU, G.; MENG, Q.; HUANG, F.; YU, D. GmNAC5, a NAC transcription factor, is a transient response regulator induced by abiotic stress in soybean. **Scientific World Journal**, v. 2013, p. 768-972, 2013.

KALEFETOGLU, T.; EKMEKÇI, Y. The effects of drought on plants and tolerance mechanisms. **G. U. Journal of Science**, v. 18, n. 4, p. 723-740, 2005.

KARIA, C. T.; ANDRADE, R. P.; CHARCHAR, M. J. D.; GOMES, A. C. Caracterização morfológica de acessos do gênero *Stylosanthes* no banco ativo de germoplasma da Embrapa Cerrados – Coleção 1994/1995. In: Boletim de desenvolvimento de pesquisa, n. 72. Planaltina: Embrapa Cerrados, 2002. 24 p.

KASUGA, M.; MIURA, S.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. A combination of the *Arabidopsis* DREB1A gene and stress-inducible *rd29a* promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. **Plant and Cell Physiology**, v. 45, n. 3, p. 346-350, 2004.

KAUR, G.; KUMAR, S.; NAYYAR, H.; UPADHYAYA, H. D. Cold stress injury during the pod-filling phase in chickpea (*Cicer arietinum* L.): effects on quantitative and qualitative components of seeds. **Journal of Agronomy and Crop Science**, v. 194, p. 457-464, 2008.

KIEFER, F.; ARNOLD, K.; KÜNZLI, M.; BORDOLI, L.; SCHWEDE, T. The SWISS-MODEL Repository and associated resources. **Nucleic Acids Research**, v. 37, p. 387-392, 2009.

KIKUCHI, K.; UEGUSHI-TANAKA, M.; YOSHIDA, K. T.; NAGATO, Y.; MATSUSOKA, M.; HIRANO, H. Y. Molecular analysis of the NAC gene family in rice. **Molecular Genomics and Genetics**, v. 262, p. 1047-1951, 2000.

KIM, S. Y.; KIM, S. G.; KIM, Y. S.; SEO, P. J.; BAE, M.; YOON, H. K.; PARK, C. M. Explorim membrane-associated NAC transcription factors in *Arabidopsis*: implications for membrane biology in genome regulation. **Nucleic Acids Research**, v. 35, n. 1, p. 203-213, 2007.

KIM, S.; KIM, S.; PARK, C. A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUST in *Arabidopsis*. **Planta**, v. 226, p. 647-654, 2007.

- KNIGHT, H.; KNIGHT, M. R. Abiotic stress signalling pathways: specificity and cross-talk. **Trends in Plant Science**, v. 6, p. 262-267, 2001.
- KOPP, J.; SCHWEDE, T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. **Nucleic Acids Research**, v. 32, p. 230-234, 2003.
- LAVIN, M.; SCHRIRE, B.; LEWIS, G.; PENNINGTON, R. T.; DELGADO-SALINAS, A.; THULIN, M.; HUGHES, C.; BEYRA MATOS, A.; WOJCIECHOWSKI, M. F. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. **Philosophical Transactions of the Royal Society**, v. 359, p. 1509-1522, 2004.
- LAWLOR, D. W. Limitation to photosynthesis in water-stressed leaves: stomata vs metabolism and the role of ATP. **Annals of Botany**, v. 89, p. 871-885, 2002.
- LEI, W.; TONG, Z.; SHENGYAN, D. Effect of drought and rewatering on photosynthetic physioecological characteristics of soybean. **Acta Ecologica Sinica**, v. 26, n. 7, p. 2073-2178, 2006.
- LEWIS, G. P.; SCHRIRE, B.; MACKINDER, B.; LOCK, M. **Legumes of the world**. Royal Botanic Gardens, Kew, p. 592, 2005.
- LIMA, A. L. S.; DAMATTA, F. M.; PINHEIRO, H. A.; TOTOLA, M. R.; LOUREIRO, M. E. Photochemical responses and oxidative stress in two clones of *Coffea canephora* under deficit conditions. **Environment and Experimental Botany**, v. 47, p. 239-247, 2002.
- LIU, T.; SONG, X.; DUAN, W.; HUANG, Z.; LIU, G.; LI, Y. Genome-wide analysis and expression patterns of NAC transcription factor family under different developmental stages and abiotic stresses in Chinese cabbage. **Plant Molecular Biology Reporter**, v. 32, p. 1041-1056, 2014.
- LIU, T.; ZHU, S.; TANG, Q.; TANG, S. Identification of 32 full-length NAC transcription factors in ramie (*Boehmeria nivea* L. Gaud) and characterization of the expression pattern of these genes. **Molecular Genetics and Genomics**, v. 289, p. 675-684, 2014.
- LPWG. THE LEGUME PHYLOGENY WORKING GROUP. Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. **Taxon**, v. 62, p. 217-248, 2013.
- MANTRI, N.; PATADE, V.; PENNA, S.; FORD, R.; PANG, E. Abiotic Stress Response in Plants: Present and Future. In: AHMAD, P.; PRASAD, M.N.V. **Abiotic Stress Response in Plants: Metabolism, Productivity and Sustainability**. Springer Science+Business Media. p. 18-36. LLC, 2012.
- MARCHLER-BAUER, A.; ANDERSON, J. B.; CHERUKURI, P. F.; DEWEESE-SCOTT, C.; GEER, L. Y.; GWADZ, M.; HE, S.; HURWITZ, D. I.; JACKSON, J. D.; KE, Z.; LANCZYCKI, C. J.; LIEBERT, C. A.; LIU, C.; LU, F.; MARCHER, G. H.; MULLOKANDOV, M.; SHOEMAKER, B. A.; SIMONYAN, V.; SONG, J. S.; THIESSEN, P. A.; YAMASHITA, R. A.; YIN, J. J.; ZHANG, D.; BRYANT, S. H. CDD: a Conserved Domain Database for protein classification. **Nucleic Acids Research**, v. 33, p. 192-196, 2005.

MARCHLER-BAUER, A.; DERBYSHIRE, M. K.; GONZALES, N. R.; LU, S.; CHITSAZ, F.; GEER, L. Y.; GEER, R. C.; HE, J.; GWARDZ, M.; HURWITZ, D. I.; LANCZYCKI, C. J.; LU, F.; MARCHLER, G. H.; SONG, J. S.; THANKI, N.; WANG, Z.; YAMASHITA, R. A.; ZHANG, D.; ZHENG, C.; BRYANT, S. H. CDD: NCBI's conserved domain database. **Nucleic Acids Research**, v. 43, 2015.

MARCHLER-BAUER, A.; PANCHENKO, A. R.; SHOEMAKER, B. A.; THIESSEN, P. A.; GEER, L. Y.; BRYANT, S. H. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. **Nucleic Acids Research**, v. 30, p. 281-283, 2002.

MARECHAL, R.; MACHERPA, J.M.; STAINER, F. Étude taxonomique d'un groupe complexe d'espèces de genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base des données morphologiques et polliniques, traitées par l'analyse informatique. **Boissiera**, v. 28, p. 1-273, 1978.

MITTLER, R. Abiotic stress, the field environment and stress combination. **Trends in Plant Science**, v. 11, n. 1, p. 15-19, 2006.

MULLER, P.; LI, X-P.; NIYOGI, K. K. Non-photochemical quenching: a response to excess light energy. **Plant Physiology**, v. 125, p. 1558-1566, 2001.

NAKASHIMA, K.; TAKASAKI, H.; UNYA MIZOI, J.; SHINOZAKI, K.; YAMAGUSHI-SHINOZAKI, K. NAC transcription factors in plant abiotic stress responses. **Biochimica et Biophysica Acta**, v. 1819, p. 97-103, 2012.

NURUZZAMAN, M.; MANIMEKALAI, R.; SHARONI, A.; SATOH, K.; KONDOH, H.; OOKA, H. Genome-wide analysis of NAC transcription factor family in rice. **Gene**, v. 465, p. 30-44, 2010.

OHASHI, Y.; NAKAYAMA, H.; SANEOKA, H.; FUJITA, K. Effects of drought stress on photosynthetic gas exchange, chlorophyll fluorescence and stem diameter of soybean plants. **Biologia Plantarum**, v. 50, p. 138-141, 2006.

OLSEN, A. N.; ERNST, H. A.; LEGGIO, L. L.; SKRIVER, K. DNA-binding specificity and molecular functions of NAC transcription factors. **Plant Science**, v. 169, p. 169-785, 2005.

OLSEN, A. N.; ERNST, H. A.; LEGGIO, L. L.; SKRIVER, K. NAC transcription factors: structurally distinct, functionally diverse. **Trends in Plant Science**, v. 10, p. 79-87, 2005.

ONOFRE, A. V. C. **Diversidade genética e avaliação de genótipos de feijão-caupi contrastantes para resistência aos estresses bióticos e abióticos com marcadores SSR, DAF e ISSR**. 2008. Dissertação de Mestrado – Universidade Federal de Pernambuco, Recife, 2008.

OOKA, H.; SATOH, K.; DOI, K.; NAGATA, T.; OTOMO, Y.; MURAKAMI, K.; MATSUBARA, K.; OSATO, N.; KAWAI, J.; CARNINCI, P.; HAYASHIZAKI, Y.; SUZUKI, K.; KOJIMA, K.; TAKAHARA, Y.; YAMAMOTO, K.; KIKUCHI, S. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. **DNA Research**, v. 10, n. 6, p. 239-247, 2003.

PADULOSI, S.; NG, N. Q. Origin, taxonomy and morphology of *Vigna unguiculata* (L.) Walp. In: SINGH, B.B.; MOHAN RAJ, D.R.; DASHIEL, K.E.; JACKAI, L.E.N. (Ed.) *Advances in cowpea research*, Tsukuba: IITA: JIRCAS, p. 1-12, 1997.

PELLEGRINESCHI, A.; REYNOLDS, M.; PACHECO, M.; BRITO, R. M.; ALMERAYA, R.; YAMAGUCHI-SHINOZAKI, K.; HOISINGTON, D. Stress-induced expression in wheat of the *Arabidopsis thaliana* DREB1A gene delays water stress symptoms under greenhouse conditions. **Genome**, v. 47, p. 493-500, 2004.

PINHEIRO, G. L.; MARQUES, C. S.; COSTA, M. D.; REIS, P. A.; CARVALHO, C. M.; FIETTO, L. G.; FONTES, E. P. Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. **Gene**, v. 444, n. 1-2, p.10-23, 2009.

POLHILL, R. M.; RAVEN, P. H.; STIRTON, C. H. Evolution and systematics of the Leguminosae. In: *Advances in Legumes Systematics 1* (R.M. Polhill & P.H. Raven, ed.). Royal Botanic Gardens, Kew, p.1-26, 1981.

RAMALHO, M. A. P.; SILVA, G. S.; DIAS, L. A. S. Genetic plant improvement and climate changes. **Crop Breeding and Applied Biotechnology**, v. 9 n. 2, p. 189-195, 2009.

REN, B.; ROBERT, F.; WYRICK, J. J.; APARICIO, O.; JENNINGS, E. G.; SIMON, I.; ZEITLINGER, J.; SCHREIBER, J.; HANNET, N.; KANIN, E.; VOLKERT, T. L.; WILSON, C. J.; BELL, S. P.; YOUNG, R. A. Genome-Wide location and function of DNA binding proteins. **Science**, v. 290, p. 2306-2309, 2000.

RIECHMANN, J. L.; HEARD, J.; MARTIN, G.; REUBER, L.; JIANG, C.; KEDDIE, J. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. **Science**, v. 290, p. 2105–2110, 2000.

RONDE, J. A. de; LAURIE, R. N.; CAETANO, T.; GREYLING, M. M.; KEREPESEI, I. Comparative study between transgenic and non-transgenic soybean lines proved transgenic lines to be more drought tolerant. **Euphytica**, v. 138, p. 123-132, 2004.

RUIZ-LOZANO, J. M. Arbuscular mycorrhizal symbiosis and alleviation of osmotic stress: new perspectives for molecular studies. **Mycorrhiza**, v. 13, p. 309-317, 2003.

RUSHTON, P. J.; BOKOWIEC, M. T.; HAN, S.; ZHANG, H.; BRANNOCK, J. F.; CHEN, X.; LAUDEMAN, T. W.; TIMKO, M. P. Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. **Plant Physiology**, v. 147, p. 280–295, 2008.

RUSHTON, P. J.; SOMSSICH, I. E. Transcriptional control of plant genes responsive to pathogens. **Current Opinion in Plant Biology**, v. 1, p. 311-315, 1998.

SAIRAN, R. K.; TYAGI, A. Physiology and molecular biology of salinity stress tolerance in plants. **Current Science**, v. 86, n. 3, p. 407-421, 2004.

SAITOH, K.; MAHMOOD, T.; KURODA, T. Effect of moisture stress at different growth stages on flowering and pod set in determinate and indeterminate soybean cultivars. **Japan Journal of Crop Science**, v. 68, p. 537-544, 1999.

SAITOU, N.; NEI, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, p. 406-425, 1987.

SALDANHA, A. J. Java Treeview – extensible visualization of microarray data. **Bioinformatics**, v. 20, n. 17, p. 3246-3248, 2004.

SALINET, L. H. **Avaliação fisiológica e agrônômica de soja geneticamente modificada para maior tolerância à seca**. 2009. Dissertação (Mestrado em Ciências) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2009.

SANTOS, R. F.; CARLESSO, R. Déficit hídrico e os processos morfológicos e fisiológicos das plantas. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 2, n. 3, p. 287-294, 1998.

SCHENK, P. M.; KAZAN, K.; WILSON, I.; ANDERSON, J. P.; RICHMOND, T.; SOMERVILLE, S. C.; MANNERS, J. M. Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. **Proceedings of the National Academy of Sciences**, v. 97, p. 11655-11660, 2000.

SCHULTZE-KRAFT, R.; COSTA, N. M. S.; FLORES, A. *Stylosanthes macrocephala* M. B. Ferr. et S. Costa: collection and preliminary agronomic evaluation of a new tropical pasture legume. **Tropical Agriculture**, v. 61, p. 229-240, 1984.

SCHWARZ, R.; DAYHOFF, M. Matrices for detecting distant relationships. In: DAYHOFF, M. **Atlas of protein sequences**. 1979. p. 353-358.

SEKI, M.; NARUSAKA, M.; ABE, H.; KASUGA, M.; YAMAGUCHI-SHINOZAKI, K.; CARNICI, P.; HAYASHIZAKI, Y.; SHINOZAKI, K. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using full-length cDNA microarray. **Plant Cell**, v. 13, p. 61-72, 2001.

SERRAJ, R.; SINCLAIR, T. R. Osmolyte accumulation: can it really help increase crop yield under drought conditions? **Plant, Cell and Environment**, v. 25, p. 333-341, 2002.

SHEN, H.; YIN, Y.; CHEN, F.; XU, Y.; DIXON, R.A. A bioinformatics analysis of NAC genes for plant cell wall development in relation to lignocellulosic bioenergy production. **Bioenergy Research**, v. 2, p. 217-232, 2009.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. **Current Opinion in Plant Biology**, v. 3, p. 217-223, 2000.

SIEVERS, F.; WILM, A.; DINEEN, D.; GIBSON, T.J.; KARPLUS, K.; LI, W.; LOPEZ, R.; MCWILLIAM, H.; REMMERT, M.; SÖDING, J.; THOMPSON, J.D.; HIGGINS, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, 539, 2011.

- SILVESTRE, M. F. **Análise de crescimento, produção e qualidade fisiológica de sementes de *Stylosanthes* spp. na baixada fluminense do Rio de Janeiro**. 2013. Dissertação de Mestrado — Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2013.
- SINGH, K. B.; FOLEY, R. C.; OÑATE-SÁNCHEZ, L. Transcription Factors in Plant Defense and Stress Responses. **Current Opinion in Plant Biology**, v. 5, p. 430-436, 2002.
- SOMOGYI, R.; GRELLER, L. D. The dynamics of molecular networks: Applications to therapeutic discovery. **Drug Discovery Today**, v. 6, n. 24, p. 1267-1277, 2001.
- SORATTO, R. P.; ORIVALDO, A. R. F.; RODRIGUES, R. A. F.; BUZETTI, S.; SILVA, T. R. B. Resposta do feijoeiro ao preparo do solo, manejo de água e parcelamento do nitrogênio. **Acta Scientiarum: Agronomy**, Maringá, v. 25, n. 1, p. 89-96, 2003.
- SOUER, E.; VAN HOUWELINGEN, A.; KLOOS, D.; MOL, J.; KOES, R. The No Apical Meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. **Cell**, v. 85, p.159-170, 1996.
- SPRENT, J.I. **Nodulation in Legumes**. Kew: The Cromwell Press, 2001. p. 146
- STACE, H. M.; Edye, L. A. (Ed.). **The biology and agronomy of *Stylosanthes***. Sidney: Academic Press, 1984.
- STILLER, I.; DULAI, S.; KONDRÁK, M.; TARNAL, R.; SZABÓ, L.; TOLDI, O.; BÁNFALVI, Z. Effects of drought on water content and photosynthetic parameters in potato plants expressing the trehalose-6-phosphate synthase gene of *Saccharomyces cerevisiae*. **Planta**, v. 227, n. 2, p. 299–308, 2008.
- TAIZ, L.; ZEIGER, E. **Plant Physiology**. Massachusetts: Publisher Sunderland, 2006. p. 705
- TAJI, T.; OHSUMI, C.; IUCHI, S.; SEKI, M.; KASUGA, M.; KOBAYASHI, M.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Importance roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. **Plant Journal**, v. 29, p. 417-426, 2002.
- TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. **Molecular Biology and Evolution**, v. 30, p. 2725-2729, 2013.
- TAOKA, K.; YANAGIMOTO, Y.; DAIMON Y.; HIBARA, K.; AIDA, M.; TASAKA, M. The NAC domain mediates functional specificity of CUP-SHAPED COTYLEDON proteins. **Plant Journal**, v. 40, p. 462-473, 2004.
- THAKUR, P.; KUMAR, S.; MALIK, J.A.; BERGER, J.D.; NAYYAR, H. Cold stress effects on reproductive development in grain crops: an overview. **Environmental and Experimental Botany**, v. 67, p. 429–443, 2010.
- THE ARABIDOPSIS INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature**, v. 408, p. 796–815, 2000.

- TRAN, L. S.; NAKASHIMA, K.; SAKUMA, Y.; SIMPSON, S. D.; PUJITA, Y.; MARUYAMA, K. Isolation and functional analysis of Arabidopsis stress inducible NAC transcription factor that bind to a drought responsive *cis*-element in the early responsive to dehydration stress 1 promoter. **Plant Cell**, v. 16, p. 2481-2498, 2004.
- UNAMBA, C.I.N.; NAG, A.; SHARMA, R.K. Next Generation Sequencing Thechnologies: The doorway to the unexplored Genomics of Non-Model Plants. **Frontiers in Plant Science**, v. 6, p. 1-16, 2015.
- VASELLATI, V.; OESTERHELD, D.; MEDAN, D.; LORETI, J. Effects of flooding and drought on the anatomy of *Paspalum dilatatum*. **Annals of Botany**, v. 88, p. 355-360, 2001.
- Vorasoot, N.; Songsri, P.; Akkasaeng, C.; Jogloy, S.; Patanothai, A. **Effect of water stress on yield and agronomic characters of peanut**. *Songklanakarin Journal of Science and Technology*, v. 25, n. 3, p. 283–288, 2003.
- WANG, W.; VINOCUR, B.; SHOSEYOV, O.; ALTMAN, A. Biotechnology of plant osmotic stress tolerance: physiological and molecular considerations. **Acta Horticulturae**, v. 560, p. 285-292, 2001.
- WATKINS, S. M.; HAMMOCK, B. D.; NEWMAN, J. W.; GERMAN, J. B. Individual metabolism should guide agriculture toward foods for improved health and nutrition. **The American Journal of Clinical Nutrition**, v. 74, n. 3, p. 283-286, 2001.
- WELNER, D. H.; LINDEMOSE, S.; GROSSMAN, J. G.; MOLLEGAARD, N. E.; OLSEN, A. N.; HELGSTRAND, C.; SKRIVER, K.; LO LOGGIO, L. DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. **Biochemical Journal**, v. 444, n. 3, p. 395-404, 2012.
- WESTBROOK, J.; FENG, Z.; CHEN, L.; YANG, H.; BERMAN, H. M. The Protein Data Bank and structural genomics. **Nuclear Acids Research**, v. 31, p. 489-491, 2003.
- WU, J.; WANG, L.; WANG, S. Comprehensive analysis and discovery of drought-related NAC transcription factors in common bean. **BMC Plant Biology**, v. 16, p. 193-205, 2016.
- XIE, Q.; FRUGIS, G.; COLGAN, D.; CHUA, N.H. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. **Genes & Development**, v. 14, p. 3024–3036, 2000
- XIE, Q.; SANZ-BURGOS, A.; GUO, H.; GARCÍA, J.; GUTIÉRREZ, C. GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. **Plant Molecular Biology**, v. 39, p. 647–656, 1999.
- XIONG, J. Introduction to Biological Databases. In: XIONG, J (ed.). **Essencial Bioinformatics**. Cambridge, UK: Cambridge University Press, 2006. p. 10-27.
- XU, B.; OHTANI, M.; YAMAGUSHI, M.; TOYOOKA, K.; WAKAZAKI, M.; SATO, M. Contribution of NAC transcription factors to plant adaptation to land. **Science**, v. 343, p. 1505-1508, 2014.

YAMAGUCHI-SHINOZAKI, K.; KASUGA, M.; LIU, Q.; NAKASHIMA, K.; SAKUMA, Y.; ABE, H.; SHIWARI, Z.K.; SEKI, M.; SHINOZAKI, K. Biological mechanisms of drought stress response. **JIRCAS Working Reports**, v. 23, p. 1-8, 2002.

YAMAGUCHI, M.; KUBO, M.; FUKUDA, H.; DEMURA, T. Vascularrelated NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in Arabidopsis roots and shoots. **The Plant Journal**, v. 55, p. 652-664, 2008.

ZHANG, M.; DUAN, L.; TIAN, X.; HE, Z.; LI, J.; WANG, B.; Li, Z. Uniconazole, induced tolerance of soybean to water deficit in relation to changes in photosynthesis, hormones and antioxidant system. **Journal of Plant Physiology**, v. 164, p. 709-717, 2007.

ZHENG, X.; TANG, S.; ZHU, S.; DAI, Q.; LIU, T. Identification of a NAC Transcription Factor Family by Deep Transcriptome Sequencing in Onion (*Allium cepa* L.). **Plos One**, v.11, n. 6, 2016.

ZHU, T.; NEVO, E.; SUN, D.; PENG, J. Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. **Evolution**, v. 66, p. 1833–1848, 2012.