

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

MITALLE KAREN DA SILVA MATOS

**Expressão Diferencial e Diversidade de Fatores de
Transcrição da Família MYB em Feijão-Caupi**

RECIFE
2015

MITALLE KAREN DA SILVA MATOS

**Expressão Diferencial e Diversidade de Fatores de
Transcrição da Família MYB em Feijão-Caupi**

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

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

Coorientadora: Dr^a Lidianne L. Barbosa Amorim

RECIFE, 2015

Catálogo na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

Matos, Mitalle Karen da Silva

Expressão diferencial e diversidade de fatores de transcrição da família MYB em feijão-caupi / Mitalle Karen da Silva Matos. – Recife: O Autor, 2015.

137 f.: il.

Orientadora: Ana Maria Benko-Iseppon, Lidiane L. Barbosa Amorim

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Ciências Biológicas. Pós-graduação em Genética, 2015.

Inclui referências e anexos

1. Genética vegetal 2. Regulação de expressão gênica I. Benko-Iseppon, Ana Maria (orient.) II. Amorim, Lidiane L. Barbosa III. Título.

581.35

CDD (22.ed.)

UFPE/CCB-2015-210

MITALLE KAREN DA SILVA MATOS

**Expressão Diferencial e Diversidade de Fatores de Transcrição
da Família MYB em Feijão-Caupi**

Aprovado em 02/03/2015

Banca Examinadora:

Dra. Ana Maria Benko-Iseppon
Universidade Federal de Pernambuco

Dr. Tercílio Calsa Júnior
Universidade Federal de Pernambuco

Dr. Péricles de Albuquerque Melo Filho
Universidade Federal Rural de Pernambuco

Dr. Luís Carlos Belarmino da Silva
Universidade Federal de Pernambuco

RECIFE, 2015

*Aos meus pais, **Paulo Jorge e Sônia Maria**
E ao meu noivo, **Luan Ramos***

Dedico.

Agradecimentos

A **Deus**, fonte de sabedoria suprema, e à **Maria Santíssima**, que, guiados por seu amor incondicional, me conduziram e me ampararam nos momentos de dificuldade desta caminhada.

Aos meus pais, **Paulo Jorge** e **Sônia Maria**, que sempre foram minha base, minha força, meu porto seguro e a quem devo tudo o que sou.

Ao meu amigo e meu amor **Luan Ramos**, por suas madrugadas dedicadas a mim, por seu cuidado, preocupação, paciência à minha falta de tempo, mas principalmente por ter sido meu parceiro e companheiro nesse final de Mestrado. Lhe sou eternamente grata.

À professora **Ana M. Benko-Iseppon**, um exemplo de mulher e pesquisadora, por toda confiança e paciência dedicados a mim e à execução deste projeto.

À **Lidiane Amorim**, carinhosamente chamada de “mãe científica” e a quem tenho grande respeito e admiração, por toda sua ajuda, conhecimento e tempo dedicados a mim e ao projeto.

Ao **João Pacífico**, por toda sua paciência em compartilhar seu conhecimento comigo, mostrando a luz aos meus desesperos e me acalmando com sua sabedoria.

À minha amiga irmã **Flávia Araújo**, por toda parceria, fidelidade, companheirismo e cumplicidade que compartilhamos nesses anos de caminhada.

Aos todos os amigos e amigas que fazem parte da família LGBV, por todos os momentos vividos, pelas boas e numerosas risadas e os valiosos conhecimentos compartilhados. Momentos únicos, com cada um, que levarei comigo sempre. Em especial, agradeço àqueles que ajudaram diretamente na

execução deste trabalho: **Neto Ferreira, Manassés Daniel, Marx de Oliveira, Uno Pacífico, Roberta Lane e Valesca Pandolfi.**

Aos meus amigos e companheiros **Vanessa Cristina, Pollyana Karla, Marcos Regueira e Santelmo Vasconcelos**, pelas inúmeras risadas que compartilhamos na ‘Casa Amarela’ e por terem feito da minha caminhada mais prazerosa e alegre morando com vocês.

Aos membros da banca examinadora, **Dr. Tercílio Calsa, Dr. Péricles de Albuquerque e Dr. Luis Carlos Belarmino** pelas contribuições para o aperfeiçoamento do trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – **CNPq**, pela bolsa concedida durante a execução deste trabalho.

À Universidade Federal de Pernambuco (**UFPE**), por meio do Programa de Pós-Graduação em Genética (**PPGG**), pela oportunidade e suporte durante todo o período do Mestrado.

MUITO OBRIGADA!!!

“Um pouco de ciência nos afasta de Deus.

Muito, nos aproxima”.

Louis Pasteur

Resumo

Danos causados por doenças virais estão entre os principais fatores limítrofes da produtividade do feijão-caupi. Em condições de estresse os fatores de transcrição (TFs) participam ativamente das etapas iniciais do processo de detecção e sinalização, regulando a expressão de vários grupos gênicos. Neste sentido, objetivou-se caracterizar TFs da família MYB e avaliar sua expressão diferencial frente à infecção viral, bem como determinar genes de referência (RGs) para normalização dos dados em RT-qPCR sob diferentes condições de estresse e controles. Por meio de análises *in silico* no banco NordEST, identificamos no transcriptoma do feijão-caupi um total de 86 candidatos a TF MYB, classificados em três subfamílias. A análise dos componentes estruturais do domínio R2R3-MYB permitiu observar a conservação dos aminoácidos característicos desta classe protéica em feijão-caupi. Por sua vez, o padrão de distribuição em pseudocromossomos de *Phaseolus vulgaris* indicou que genes MYB sofreram duplicações em *tandem* e intercromossomais, contribuindo para sua expansão no feijão-caupi. A análise filogenética formou 18 subclados, apoiados pela estrutura dos motivos funcionais da região C-terminal das proteínas. Das tags SuperSAGE diferencialmente expressas sob infecção viral, três foram reguladas positivamente, indicando a participação de candidatos MYB na resposta ao estresse viral. Dos sete RGs avaliados em três conjuntos experimentais, β -tubulina, Skip16 e Act2/7 + Skip16 foram as melhores combinações para seca, salinidade e vírus, respectivamente, podendo ser recomendados como normalizadores para estudos de expressão diferencial em feijão-caupi. Neste estudo identificamos a maior família de TFs em plantas observando sua participação ativa na resposta de defesa contra estresses em feijão-caupi.

Palavras-chave: Bioinformática; Perfil de Expressão Gênico; *Vigna unguiculata*; Estresse ambiental; Genes *Housekeeping*.

Abstract

Damages caused by viral diseases are among the main factors affecting the cowpea productivity. Under stress conditions, transcription factors (TFs) actively participate in the initial stages of the detection and signaling process by regulating the expression of various gene groups. In this sense, the objective of the present work was to characterize members of the MYB TF-family and evaluate their differential expression under viral infection, also determining reference genes (RGs) for data normalization in RT-qPCR under different stress and control conditions. Using *in silico* approaches to analyze the NordEST databank, a total of 86 MYB TF-candidates could be identified, being classified into three subfamilies. An analysis of the structural components of the R2R3-MYB domain allowed the identification of conserved amino acid residues of this protein class in cowpea. In turn, the MYB distribution pattern in the pseudochromosomes of *Phaseolus vulgaris* indicated that MYB members suffered in tandem and interchromosomal duplications, contributing to their expansion in cowpea. Phylogenetic analysis formed 18 subclades, supported by structural features of motifs in the C-terminal region of the protein. Of differentially expressed SuperSAGE tags under viral infection, three were upregulated, indicating the involvement of MYB candidates in response to viral stress. Considering the seven tested RGs under three experimental conditions, β -tubulin, Skip16 and Act2/7 + Skip16 were the best combinations for drought, salinity and viruses, respectively, recommended as normalizer genes in studies of differential expression in cowpea. In the present work we identified members of the largest family of TFs in plants observing their active participation in defense against stress response in cowpea.

Key words: Bioinformatics; Gene Expression Profiling; *Vigna unguiculata*; Environmental stress; Housekeeping genes.

Lista de Figuras

Revisão da Literatura

Figura 1: Modelo de uma proteína R2R3-MYB. **(A)** Estrutura tridimensional da proteína c-MYB de um vertebrado ligada ao DNA. A primeira, segunda e terceira hélices de cada repetição estão destacadas em vermelho, amarelo e azul, respectivamente e o DNA é apresentado entre as duas repetições. **(B)** Representação linear de um domínio R2R3-MYB encontrado em plantas (Fonte: Heine, 2006) **25**

Figura 2: Representação da estrutura primária e secundária das quatro subfamílias do TF MYB presente nos vegetais. (H) hélice; (T) volta; (W) triptofano; (X) aminoácido (Fonte: Dubos *et al.*, 2010) **26**

Figura 3: Fluxograma da técnica SuperSAGE. (Fonte: Traduzido de Matsumura *et al.*, 2008) **39**

Figura 4: Construção de uma biblioteca de cDNA. Os RNAs longos primeiramente são convertidos em uma biblioteca de fragmentos de cDNA (por meio da fragmentação do RNA ou do DNA). Subsequentemente, os adaptadores de sequenciamento (azul) são adicionados a cada fragmento de cDNA e uma sequência curta é obtida utilizando tecnologia de sequenciamento de alto rendimento. As *reads* resultantes são alinhadas com o genoma ou transcriptoma de referência e classificadas em três tipos: *reads* em exon, *reads* de junção e *reads* na cauda poli (A). (Fonte: Traduzido de Wang *et al.*, 2009b) **41**

Capítulo I

Figura 1: Alinhamento da sequência de aminoácidos de 59 genes com domínio R2R3-MYB do feijão-caupi utilizando o Clustal Omega. As setas indicam os aminoácidos conservados em todos os genes e os asteriscos indicam os resíduos de Triptofano que formam o centro hidrofóbico da proteína. Gráficos na base da figura indicam a conservação e a qualidade do alinhamento **57**

Figura 2: Localizações cromossômicas e regiões de duplicação das proteínas MYB do feijão-caupi. O número do cromossomo é indicado na região externa e a posição dos seus centrômeros por uma linha branca. Linhas coloridas indicam supostas duplicações das sequências MYB; **(A)** Distribuição e duplicações das proteínas pertencentes às subclasses R2R3-MYB e **(B)** R1-MYB e R3-MYB **59**

Figura 3: Dendrograma refletindo padrões da estrutura secundária dos motivos conservados na região C-terminal das proteínas R2R3-MYB e denominações dos subgrupos das proteínas R2R3-MYB do feijão-caupi (Vu). **(A)** A árvore NJ inclui 64 proteínas R2R3-MYB do feijão-caupi e apresenta 18 subgrupos (C1-C18) marcados com fundos coloridos, para facilitar a identificação da subfamília com alto valor preditivo. Os números ao lado dos ramos representam valores de bootstrap ($\geq 60\%$) com base em 1000 repetições. **(B)** Estrutura secundária dos motivos conservados na região C-terminal das proteínas R2R3-MYB em 18 subgrupos. Os motivos foram detectados utilizando o software MEME e são representados graficamente como caixas desenhadas à escala para uma proteína MYB representante de cada subfamília. Caixas em azul representam o domínio R2R3. *Proteínas VuMYB que não apresentam o padrão de motivo característico do seu subgrupo

61

Figura 4: Quantificação relativa dos genes MYB sob diferentes tempos de estresse viral; C30T30 (controle tempo 30 e tratamento com vírus tempo 30); C30T60 (controle tempo 30 e tratamento com vírus tempo 60) e C16T16 (controle tempo 16 horas e tratamento com vírus tempo 16 horas). * Genes up-regulados em comparação ao controle

67

Capítulo II

Figure 1. The expression stability of seven reference genes was evaluated by geNorm program, which calculates an average expression stability values (M) based on the average pair-wise variation existing between all pairs of candidate genes for abiotic stress and six pairs for biotic stress. The lower M value indicates a more stable expression. (a) Drought stress, (b) Salinity stress and (c) Virus stress. NormFinder ranking of expression stability of seven reference genes to abiotic stress and six genes to biotic stress according to their expression variation between inter and intra groups. The lower stability values indicate a more stable expression as a reference gene. The best combination is indicated by orange color column. (d) Drought stress, (e) Salinity stress and (f) Virus stress

89

Figure 2. Evaluation of the optimum number of reference genes according to geNorm software. Pairwise variation (V) of the reference genes in **(a)** drought stress, **(b)** salinity stress, and **(c)** virus stress

92

Figure 3: RT-qPCR of abiotic and biotic stress-related genes measured at the appropriate sample time using REST2009 software. The relative expression levels of genes LTP (blue bar), CHI (red bar), and CHS (green bar) were determined after normalization. The drought treatment samples (A) were normalized by geNorm with Act2/7 and Ubq. The salinity treatment samples (B) were normalized with Ukn1 and Ubq

95

Lista de Tabelas

Capítulo I

Tabela 1: Funções designadas aos subgrupos correspondentes do agrupamento com as sequências de AtMYB, segundo a classificação designada por Dubos <i>et al.</i> (2010)	63
Tabela 2: Sequências consensos dos motivos da região C-terminal de proteínas R2R3-MYB do feijão-caupi e <i>Arabidopsis</i> identificados através de análise na ferramenta MEME	64
Tabela 3: Tags selecionadas com suas respectivas ESTs utilizadas como <i>input</i> para o desenho e amplificação dos <i>primers</i> via RT-qPCR	66

Capítulo II

Table 1. Gene name, slopes, amplification efficiency (<i>E</i>) and correlation efficiency (R^2) of each reference gene analyzed in this study	86
Table 2. Putative stress marker genes retrieved from the HT-SuperSAGE and EST analysis	93
Table 3. Primer sequences, PCR amplicon length of the selected reference genes and stress responsive genes in cowpea	99

Lista de Abreviaturas

Item	Definição
aa	Aminoácidos
ABA	<i>Abscisic Acid</i> (Ácido abscísico)
AP2/ERF	APETALA 2/Ethylene Response Factor
Arg	Arginina
AS	<i>Salicylic Acid</i> (Ácido salicílico)
Avr	<i>Avirulence</i> (Avirulência)
BLAST	<i>Basic Local Alignment Search Tool</i> (Ferramenta de Busca por Alinhamento Local)
bZIP	<i>Basic Leucine Zipper</i> (Zíper de Leucina Básica)
C1	lôcus COLORED1
CABMV	<i>Cowpea Aphid-Borne Mosaic Virus</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CD	<i>Conserved Domain</i> (Domínio Conservado)
cDNA	<i>Complementary DNA</i> (DNA complementar)
CGKB	<i>Cowpea Genomics Knowledge Base</i> (Base de Conhecimentos Genômicos de Feijão-Caupi)
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPSMV	<i>Cowpea severe mosaic virus</i> (Vírus do Mosaico Severo do Caupi)
Cys	Cisteína
DNA	<i>Deoxyribonucleic Acid</i> (Ácido Desoxirribonucleico)
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EST	<i>Expressed Sequence Tag</i> (Etiqueta de Sequência Expressa)
FACEPE	Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco
GA3	<i>Gibberellic Acid</i> (Ácido Giberélico)
GENOSOJA	Consórcio Nacional para Estudos do Genoma da Soja
GSS	<i>Cowpea Genespace Sequences</i>
HR	<i>Hypersensitive Response</i> (Resposta de Hipersensibilidade)
HTH	<i>Helix–Turn–Helix</i> (hélice-volta-hélice)

Ile	Isoleucina
IPA	Instituto Agrônômico de Pernambuco
JA	<i>Jasmonic acid</i> (Ácido Jasmônico)
JAZ	<i>Jasmonate ZIM Domain</i> (Proteínas contendo domínio jasmonato ZIM)
Leu	Leucina
MBS	<i>MYB-Binding Site</i> (Sítio de ligação MYB)
MCT	Ministério de Ciências e Tecnologia
MYB	<i>Myeloblastosis</i> (Mieloblastose)
MYC	<i>Myelocytomatosis</i> (Mielocitomastose)
NAC	NAM/ATAF1/CUC2
NCBI	<i>National Center for Biotechnology Information</i> (Centro Nacional para Informação Biotecnológica)
NLS	<i>Nuclear Localization Signal</i> (Sinal de Localização Nuclear)
NordEST	Rede Nordeste de Biotecnologia - bando de dados do transcriptoma do feijão-caupi
ORF	<i>Open Reading Frame</i> (Quadro de Leitura Aberto)
ORF-finder	<i>Open Reading Frame Finder</i> (Identificador de Quadro de Leitura Aberto)
PAMPs	<i>Pathogen-Associated Molecular Patterns</i> (Padrão Molecular Associado a Patógenos)
pb	Pares de bases
PCR	<i>Polymerase Chain Reaction</i> (Reação em cadeia da polimerase)
Phe	Fenilalanina
PR	<i>Pathogenesis Related</i> (Proteínas Relacionadas à Patogênese)
RENORBIO	Rede Nordeste de Biotecnologia
RG	<i>Resistance gene</i> (Genes de Resistência)
RNA	<i>Ribonucleic Acid</i> (Ácido ribonucleico)
mRNA	<i>messenger RNA</i> (RNAs mensageiros)
RNA-seq	<i>RNA Sequencing</i> (Sequenciamento de RNA)
RT-qPCR	<i>Real Time Quantitative PCR</i> (PCR quantitativa em Tempo Real)
SAGE	<i>Serial Analysis of Gene Expression</i> (Análise Serial da Expressão Gênica)

SAR	<i>Systemic Acquired Resistance</i> (Resistência Sistêmica Adquirida)
Tag	<i>Target</i> (Etiquetas)
TFs	<i>Transcription Factors</i> (Fatores de Transcrição)
TM	<i>Melting Temperature</i> (Temperatura de Dissociação)
Trp	Triptofano
WRKY	Sequência de aminoácido <u>WRKY</u> GQK
WTSS	<i>Whole Transcriptoma Shotgun Sequencing</i>

Sumário

1. Introdução.....	16
2. Revisão da Literatura.....	18
2.1. Mecanismos de respostas vegetais frente a estresses ambientais	18
2.1.1. Fatores de transcrição envolvidos na resposta a estresses.....	20
2.1.2. Família MYB: Estrutura, evolução e diversidade de funções.....	24
2.2. A família Fabaceae e sua importância.....	29
2.2.1. A cultura do feijão-caupi (<i>Vigna unguiculata</i>).....	31
2.3. Genômica funcional do feijão-caupi: Banco de dados NordEST.....	33
2.4. Análise global da expressão gênica.....	34
2.4.1. ESTs – <i>Expressed Sequence Tags</i>	36
2.4.2. SuperSAGE	37
2.4.3. RNA-seq	40
2.4.4. RT-qPCR	42
3. Objetivos	44
3.1 Geral	44
3.2 Específicos	44
4. Capítulo I.....	45
5. Capítulo II.....	81
6. Discussão geral.....	106
7. Conclusões gerais	108
8. Referências Bibliográficas	109
9. Anexos	117
10. Curriculum vitae (Lattes)	136

1. Introdução

As plantas têm sua expressão gênica regulada em vários níveis. Um dos principais meios de controle da expressão é realizado em nível transcricional, garantindo um menor gasto energético da célula e níveis adequados de proteínas produzidas. Esta regulação é coordenada por vários mecanismos que envolvem, dentre outros fatores, a metilação do DNA, a organização da cromatina, a ligação de sequências específicas ao DNA, os quais podem ser controlados pelos fatores de transcrição (TFs - *Transcription Factors*), entre outros. A regulação gênica controlada por diferentes famílias de TFs desempenha um papel importante, havendo convergência de vias de sinalização distintas que regulam a expressão do conjunto de genes responsivos a cada sinalização a fim de permitir uma resposta adequada (Feller *et al.*, 2011).

A família MYB (myeloblastosis) de fator de transcrição, composta por sequências contendo de uma a quatro repetições HTH (*Helix-Turn-Helix*), é uma das mais numerosas e diversificadas funcionalmente em plantas. Sua classificação é baseada no número de repetições MYB adjacentes, podendo conter um, dois, três ou quatro domínios (R1-MYB, R2R3-MYB, R3-MYB e 4R-MYB, respectivamente) (Stracke *et al.*, 2001; Du *et al.*, 2012). Diversos estudos envolvendo essa família de TFs têm sido conduzidos nos últimos anos, tendo sido constatada uma regulação tanto positiva quanto negativa mediada por esses TFs sobre a expressão de genes responsivos ao frio, salinidade, seca, patógenos, entre outros (Oh *et al.*, 2011; Shin *et al.*, 2011; Xie *et al.*, 2014).

A cultura do feijão-caupi [*Vigna unguiculata* (L.) Walp], configura-se como fonte de emprego e renda nas regiões Norte e Nordeste do Brasil, sendo uma leguminosa com grande plasticidade fisiológica, variabilidade genética e boa adaptação a diversos ambientes (Freire-Filho *et al.*, 2011). Apesar de haver acessos com características agronômicas requeridas pelo mercado consumidor,

muitos genótipos não apresentam resistência às principais doenças da cultura. Dadas as dificuldades de controle e sua forma de disseminação, o ataque de viroses à cultura do feijão-caupi constitui-se como um dos principais fatores limítrofes da produção na América Latina. No Brasil, o melhoramento do feijão-caupi baseia-se principalmente em técnicas convencionais, havendo poucos estudos associados às técnicas moleculares modernas, supondo-se que tecnologias de genômicas e transcriptômicas possam ajudar na superação dessas adversidades.

Neste contexto, o MCT (Ministério de Ciências e Tecnologia) através da Rede Nordestina de Biotecnologia (RENORBIO) apoiou um estudo funcional, estrutural e comparativa do feijão-caupi através da Rede Brasileira de Genômica do Feijão-caupi (NordEST) a qual gerou mais de 500 milhões de transcritos expressos sob diferentes condições de estresses biótico (vírus do mosaico severo do caupi e vírus transmitido por afídeos) e abiótico (seca e salinidade), tendo sido identificados potenciais candidatos dentre os fatores de transcrição e genes reguladores de cascatas de sinais, entre outros.

Dado o grande tamanho da família MYB e sua natureza funcionalmente diversa, a mineração de dados desta família de genes no transcriptoma do feijão-caupi é crucial para compreender os papéis dessa família de TF em processos biológicos dessa cultura. Além disso, a análise das relações estruturais entre proteínas MYB do feijão-caupi com ortólogos facilitará a previsão das funções de genes ainda não caracterizados.

2. Revisão da Literatura

2.1. Mecanismos de respostas vegetais frente a estresses ambientais

As plantas seguiram um curso evolutivo para viver em ambientes onde frequentemente lhes são impostas condições adversas (Velázquez *et al.*, 2011). A natureza de tais adversidades pode ser de caráter biótico (como vírus, bactérias, fungos, herbívoros e nematóides) ou abiótico (como temperaturas altas, deficiência mineral, toxicidade, déficit hídrico e elevada concentração salina no solo) (Alves *et al.*, 2013). Em função da sua condição sésil, as plantas desenvolveram ao longo da sua evolução, mecanismos de respostas de defesa que lhes permitiram detectar as mudanças ambientais, responder às condições estressantes e retornar ao seu metabolismo normal, minimizando desta forma as perdas sofridas, ao manter recursos para o crescimento e reprodução (Pinto *et al.*, 2011; Velázquez *et al.*, 2011; Atkinson e Urwin, 2012).

As respostas vegetais aos estresses são altamente complexas e envolvem mudanças nos níveis fisiológico, bioquímico e molecular, ativando um programa específico de expressão gênica referente à condição ambiental que lhes está sendo imposta (Atkinson e Urwin, 2012). Os produtos destes genes estresse-induzidos classificam-se em dois grupos: o primeiro inclui barreiras físicas, bem como as defesas pré-programadas e as patógeno-específicas, enquanto que o segundo grupo inclui proteínas reguladoras, como as quinases ou fosfatases e os fatores de transcrição (Alves *et al.*, 2013).

As barreiras físicas são estruturais e pré-formadas, limitando a invasão e infecção do patógeno. Como exemplo, a epiderme, espinhos e outras estruturas de superfície injuriosa; parede celular, que pode estar associada a uma variedade

de defesas químicas; a lignina, que dá rigidez à célula; cutina, suberina e ceras, que são substâncias gordurosas; células denominadas idioblastos, que contêm produtos químicos tóxicos ou cristais cortantes que rasgam as mandíbulas de herbívoros; bem como a formação de calos após o ataque de um micro-organismo (Freeman e Beattie, 2008).

Além de barreiras pré-formadas, as plantas têm a capacidade de reconhecer patógenos potencialmente perigosos e ativar respostas específicas de defesa pré-programada. A resistência basal ou imunidade inata é a segunda estratégia desenvolvida pelas plantas para se protegerem das doenças. Tal mecanismo é acionado pelo reconhecimento dos padrões moleculares associados a patógenos (PAMPs – *Pathogen-Associated Molecular Patterns*), ativando vias de transdução de sinal para a produção de metabólitos secundários, compostos químicos tóxicos e enzimas hidrolíticas, as quais atuam somente na presença de água (Freeman e Beattie, 2008; Krishania *et al.*, 2013).

As plantas possuem ainda outra linha de defesa mais específica contra os patógenos que resistem à imunidade inata, a chamada Reação de Hipersensibilidade (*Hypersensitive Response* - HR). A HR caracteriza-se pela interação de produtos dos genes R (*Resistance*) das plantas com o produto dos genes Avr (*Avirulence*) do patógeno resultando na morte celular localizada, além de desencadear uma cascata de sinais responsáveis pela ativação da Resistência Sistêmica Adquirida (SAR - *Systemic Acquired Resistance*). A ativação da SAR resulta na produção de metabólitos, fitoalexinas e proteínas PR (*Pathogenesis Related*) que, junto com os genes R, representam os principais mecanismos de defesa das plantas (Wanderley-Nogueira *et al.*, 2007; Krishania *et al.*, 2013).

Uma diversidade de mecanismos responsivos aos estresses evoluiu nos organismos vegetais. A ativação dos genes envolvidos nesses mecanismos é minuciosamente controlada por um grande número de proteínas reguladoras, as quais constituem a segunda estratégia de defesa e podem atuar em dois níveis da expressão gênica: **(1)** a nível transcricional (quantidade de mRNA produzido) e/ou **(2)** a nível traducional (quantidade de proteínas produzidas) (Alves *et al.*, 2014). Para a maioria dos genes, essa regulação é realizada em nível transcricional, garantindo desta forma um menor gasto energético da célula, além de níveis adequados de proteínas. Dentre os mecanismos regulatórios transcricionais podem-se citar a remodelagem da cromatina, a organização dos nucleossomos, a metilação do DNA e as interações das proteínas chamadas fatores de transcrição (*Transcription Factors* – TFs) (Feller *et al.*, 2011). De forma genérica, a regulação da transcrição é o resultado dos efeitos combinados das propriedades estruturais do DNA e suas interações com os TFs.

2.1.1. Fatores de transcrição envolvidos na resposta a estresses

Os TFs estão entre os promotores de tolerância ao estresse, pois participam de cascatas de eventos moleculares alterando diretamente a expressão de grande quantidade de genes de defesa. A importância desses fatores para a regulação dos genes estresse-induzidos se reflete na composição genômica dos vegetais. Segundo Udvardi *et al.* (2007), as plantas destinam em média 7% de seus genes para codificar TFs, já tendo sido observado na soja cultivada (*Glycine max* (L.) Merr.) que 12% do total de genes são codificadores de TFs (Wang *et al.*, 2010), comprovando assim a complexidade da regulação da transcrição gênica nos vegetais. No organismo modelo *Arabidopsis thaliana* (L.)

Heynh, um total de 27.416 genes são codificadores de proteínas, dos quais mais de 1.700 genes (6%) codificam TFs (Feller *et al.*, 2011).

TFs são proteínas que reconhecem motivos de DNA em regiões reguladoras ou intensificadoras de genes e desta forma facilitam ou inibem o acesso da RNA polimerases ao molde de DNA. Essas proteínas são compostas por, no mínimo, quatro domínios discretos: o domínio de ligação ao DNA, o sinal de localização nuclear (NLS - *Nuclear Localization Signal*), o domínio de ativação da transcrição, e o local de oligomerização, que juntos atuam como reguladores da expressão de genes-alvo por meio de transdução de sinais nas diversas vias (Du *et al.*, 2009). Além disso, eles ainda interagem com a maquinaria de transcrição gênica, com as proteínas que remodelam a cromatina e até mesmo com outros TFs. Essas proteínas atuam como reguladores-chave de inúmeros processos celulares e apresentam-se como excelentes candidatos para modificar caracteres complexos em plantas cultivadas, além de mostrarem-se como prováveis recursos tecnológicos para a próxima geração de cultivos biotecnológicos (Ambawat *et al.*, 2013).

TFs são classificados em famílias distintas de acordo com suas características estruturais e a conservação dos seus domínios de ligação ao DNA, podendo ainda essas famílias subdividir-se de acordo com o número e a distância das regiões conservadas. Exemplos incluem hélice-alça-hélice, dedos de zinco, hélice-volta-hélice e zíper de leucina (Liu *et al.*, 1999; Cai *et al.*, 2012). A modulação da função de TFs através das suas interações com outras proteínas reguladoras é um processo crucial para a ativação ou repressão de vias de transdução de sinal (como HR, o reconhecimento dos PAMPs, SAR, etc.), que

são finamente reguladas pelas interações entre proteínas com diferentes TFs (Alves *et al.*, 2014).

No decorrer dos últimos anos têm-se identificado um número substancial de proteínas interagindo com TFs envolvidos na defesa vegetal contra os agentes patogênicos. Várias famílias de TFs em planta desempenham papéis importantes na transdução de sinais frente a estresses bióticos para alterar a expressão gênica. Dentre as principais, estão a WRKY (contém a sequência de aminoácido WRKYGQK); AP2/ERF (APETALA 2/Ethylene Response Factor); bZIP (Basic Leucine Zipper); MYC (Myelocytomatosis Related Proteins); NAC (NAM/ATAF1/CUC2) e MYB (Myeloblastosis Related Proteins) (Alves *et al.*, 2014).

Os TFs da superfamília WRKY, assim chamados devido a região invariável (WRKYGQK) do domínio de ligação ao DNA, foram descritos com destaque à modulação de sua expressão gênica em plantas no processo de resposta a infecções virais, bacterianas e fúngicas (Ülker e Somssich, 2004). As interações dessas proteínas com outras moléculas podem desempenhar papéis na sinalização, transcrição, remodelação da cromatina e outros processos celulares importantes para a ativação da imunidade da planta (Chi *et al.*, 2013).

Por sua vez, TFs AP2/ERF têm sua ação regulada de três modos distintos: atuam frequentemente de forma sinérgica com outros TFs para aumentar a expressão dos genes de defesa da planta; têm sua função regulatória ativada por meio da fosforilação promovida por proteínas quinases; ou são dissociados de outras proteínas como resposta à infecção patogênica, podendo então ser transferidos ao núcleo, promovendo a expressão dos genes PR (Alves *et al.*, 2014).

A família de TFs que contém o domínio bZIP apresenta-se como uma das maiores nos eucariotos. Em plantas, essa família foi descrita atuando na regulação de genes responsivos ao estresse abiótico, à maturação das sementes, ao desenvolvimento floral e à defesa contra patógenos. As proteínas bZIP de *Arabidopsis* foram classificadas em dez grupos (A, B, C, D, E, F, G, H, I e S), sendo os membros pertencentes ao grupo D ligados com as respostas ao estresse biótico (Jakoby *et al.*, 2002).

A família MYC de TF atua regulando positivamente a expressão dos genes responsivos a lesões em plantas e negativamente aos genes de resistência do patógeno, como resposta associada à via do jasmonato (JA) produzido pelas plantas sob ataque de patógenos e herbívoros. A ativação dos TFs MYC também resulta na expressão de outras famílias de TFs, tais como MYBs, WRKYs e AP2/ERFs, bem como de genes da biossíntese do ácido jasmônico (*Jasmonic acid* - JA) e proteínas JAZ (*Jasmonate Zim Domain*), que são importantes na cascata de sinalização de defesa contra o estresse (Yan *et al.*, 2013).

Os TF NAC compreendem uma família recém-descoberta e específica dos vegetais, tendo sido observada sua expressão em diferentes tecidos e estádios de desenvolvimento (Voitsik *et al.*, 2013). Atuando nas respostas de defesa da planta contra patógenos, a primeira evidência veio do gene StNAC da batata, induzido por um ataque patogênico (Collinge e Boller, 2001). Por conseguinte, vários estudos verificaram a indução da expressão destas proteínas frente a estresses bióticos (Collinge e Boller, 2001; Oh *et al.*, 2005; Nakashima *et al.*, 2007; Wang *et al.*, 2009a).

Genes MYB constituem a maior família de TFs descrita até hoje em eucariotos. O grande tamanho dessa família nos organismos vegetais denota sua

importância no controle dos processos específicos (Ambawat *et al.*, 2013). Considerando a subdivisão da família de acordo com suas características estruturais, a subfamília R2R3-MYB é exclusiva das plantas, onde diversas proteínas têm sido caracterizadas por abordagens genéticas, verificando-se seu envolvimento no controle de processos específicos de plantas, dentre os quais, a resposta aos estresses bióticos merece destaque (Dubos *et al.*, 2010; Alves *et al.*, 2014).

2.1.2. Família MYB: Estrutura, evolução e diversidade de funções

Assim como outras famílias de TFs, as proteínas MYB apresentam duas regiões distintas: uma formada pelo domínio de ligação ao DNA altamente conservado localizado na porção N-terminal e outra região diversificada responsável pela atividade reguladora da proteína localizada na porção C-terminal (Ambawat *et al.*, 2013). O domínio MYB é geralmente composto de uma a quatro repetições (R) imperfeitas que codificam três α -hélices, onde cada repetição possui cerca de 50 aminoácidos (aa) com resíduos de Triptofano (Trp) regularmente espaçados, desempenhando função no dobramento do centro hidrofóbico do domínio. A segunda e terceira α -hélices conformam-se numa estrutura de hélice-volta-hélice (*helix–turn–helix* – HTH) formando um núcleo hidrofóbico que se liga à sequência alvo no promotor. A terceira hélice é definida como a hélice de reconhecimento e está em contato direto com o DNA (Figura 1) (Stracke *et al.*, 2001; Du *et al.*, 2009; 2012; Tombuloglu *et al.*, 2013). A sequência alvo do domínio MYB na região promotora do DNA é um motivo consenso denominado sítio de ligação ao MYB (MBS; *MYB-Binding Site*), o qual pode se apresentar de três tipos: MBS1, MBS2 e MBS2G (Kamiya *et al.*, 2002).

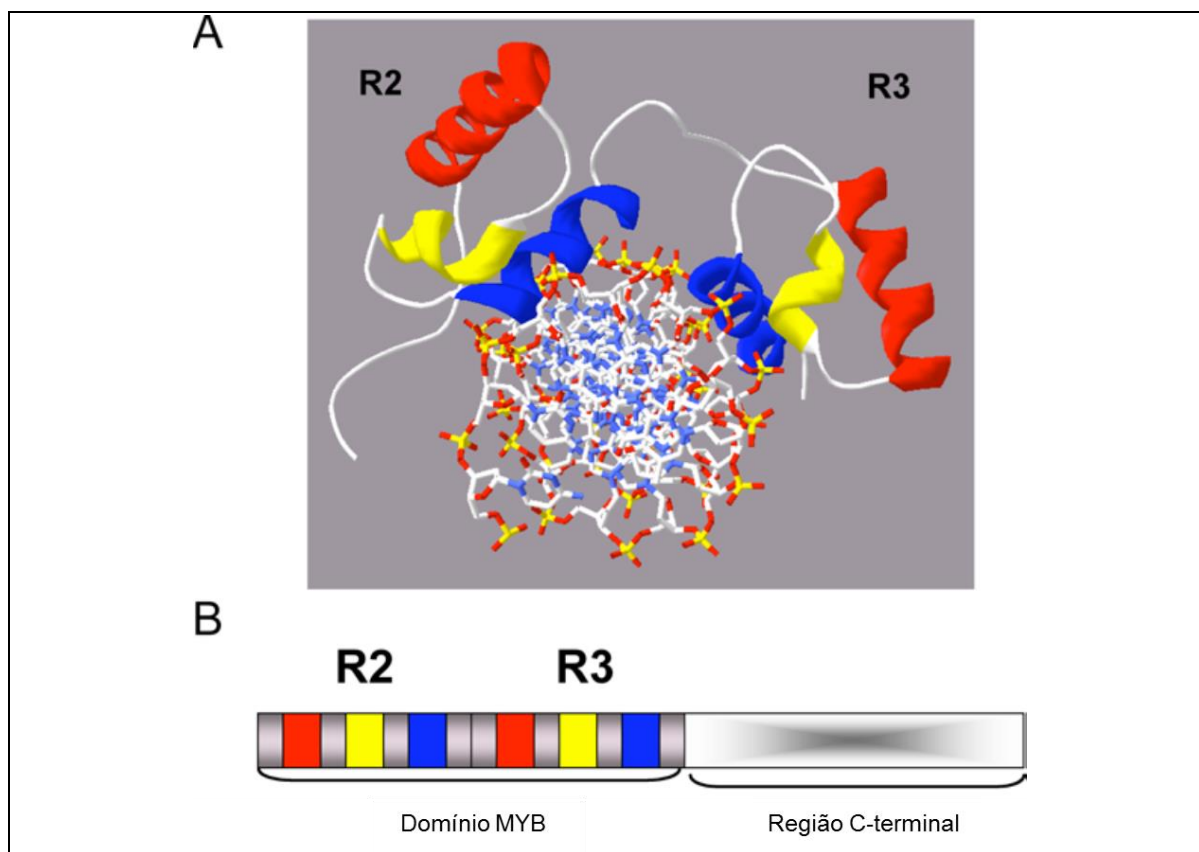


Figura 1: Modelo de uma proteína R2R3-MYB. **(A)** Estrutura tridimensional da proteína c-MYB de um vertebrado ligada ao DNA. A primeira, segunda e terceira hélices de cada repetição estão destacadas em vermelho, amarelo e azul, respectivamente e o DNA é apresentado entre as duas repetições. **(B)** Representação linear de um domínio R2R3-MYB encontrado em plantas (Fonte: Heine, 2006).

Com base no número de repetições MYB adjacentes, a família pode subdividir-se em quatro classes contendo um, dois, três ou quatro domínios (R1-MYB, R2R3-MYB, R3-MYB e R4-MYB, respectivamente - Figura 2) (Stracke *et al.*, 2001; Dubos *et al.*, 2010). Genes R2R3-MYB são específicos de plantas e constituem-se como a maior subfamília destes organismos. Ainda assim, todas as quatro classes são encontradas nos vegetais, representando o táxon com a maior diversidade das proteínas MYB (Stracke *et al.*, 2001; Du *et al.*, 2012; Ambawat *et al.*, 2013). A segunda classe mais encontrada nas plantas compreende as proteínas com uma repetição MYB completa ou parcial, designadas como "MYB-

relacionada", que se dividem em várias subclasses. A classe contendo três repetições MYB, normalmente é codificada por cinco genes em genomas de plantas superiores, sendo também encontrados na maioria dos genomas eucarióticos. Pouco se sabe sobre a classe 4R-MYB, sendo verificado que uma única proteína é codificada em vários genomas de plantas (Dubos *et al.*, 2010).

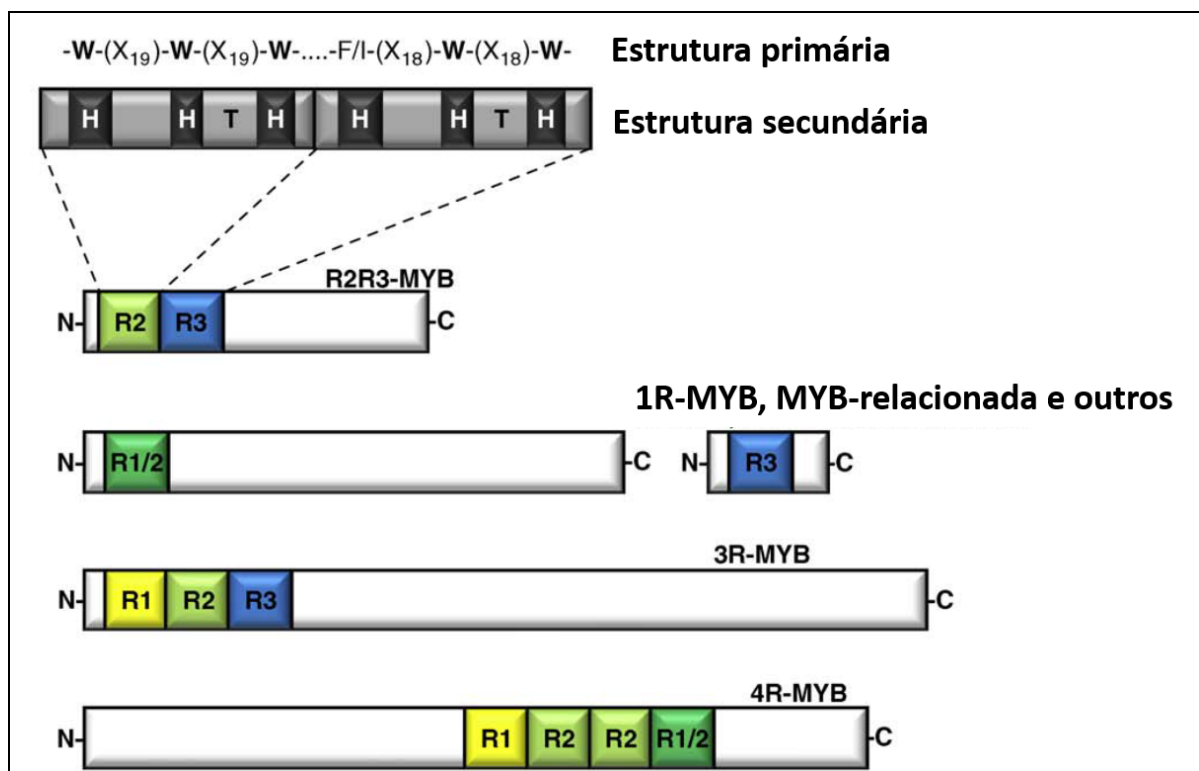


Figura 2: Representação da estrutura primária e secundária das quatro subfamílias do TF MYB presente nos vegetais. (H) hélice; (T) volta; (W) triptofano; (X) aminoácido (Fonte: Dubos *et al.*, 2010).

O primeiro gene codificador de uma proteína contendo o domínio MYB foi descrito a partir do genoma do vírus mieloblastose aviária (então denominada v-MYB), também responsável pelo nome 'MYB' do domínio (Klempnauer *et al.*, 1982). Subsequentemente, os genes c-MYB, A-MYB e B-MYB foram identificados em vertebrados e a posteriori em todos eucariotos estudados até então (Lipsick, 1996; Weston 1998). Em plantas, o primeiro gene identificado codificando uma

proteína da família MYB foi localizado no locus COLORED1 (C1), responsável por codificar um TF necessário para a síntese de antocianinas no aleurona da semente do milho (*Zea mays*) (Dubos *et al.*, 2010).

Estudos evolutivos com base nas sequências de domínios MYB de vários organismos indicam que os genes precursores de MYB nos vegetais podem ter-se originado de um ancestral que hoje é representado pelo gene B-MYB de três repetições (R3) (Ambawat *et al.*, 2013). O modelo evolutivo proposto por Lipsick (1996) sugere que houve perda da primeira repetição (R1) a partir dessas proteínas, originando a subfamília R2R3-MYB em plantas. No entanto, outra proposta evolutiva para a origem de R2R3-MYB sugere que esta subfamília tenha surgido a partir de um evento de duplicação com subsequente ganho do domínio R1, indicando que R2R3-MYB seja um precursor de R3-MYB (Jiang *et al.*, 2004).

Houveram ainda outras modificações no domínio protéico para formar o domínio R2R3-MYB específico de plantas. Esses eventos evolutivos envolveram: **(i)** a substituição do primeiro Trp da repetição R3 por outro aa hidrofóbico; **(ii)** a inserção de um resíduo de Leucina (Leu) na posição 46 entre a segunda e a terceira hélice de R2; **(iii)** a inserção de um resíduo de Cisteína (Cys) na posição 49; que tem permanecido conservada entre todos os domínios R2R3-MYB conhecidos (Dias *et al.*, 2003). Essas proteínas passaram por uma extensa amplificação no período de origem das plantas terrestres, há cerca de 500 milhões de anos, antecedendo a divisão entre mono e dicotiledôneas. Observações em vários membros de R2R3-MYB com funções específicas nos vegetais sugerem que a expansão da família tenha sido uma resposta à seleção do novo estilo de vida sésstil das plantas (Feller *et al.*, 2011).

As funções das proteínas MYB foram analisadas em diversas espécies de plantas utilizando-se análises genéticas e moleculares. Isso tem gerado uma imensa quantidade de dados acerca do papel dos TFs MYB nos vegetais e confirmado o envolvimento dessas proteínas numa diversidade de funções biológicas específicas das plantas (Ambawat *et al.*, 2013).

Vários genes R2R3-MYB foram descritos atuando de forma geral no desenvolvimento da planta (Wang *et al.*, 2008; Brownfield *et al.*, 2009; Cheng *et al.*, 2009), controlando a proliferação e diferenciação celular (Li *et al.*, 2009; Zhang *et al.*, 2009; Dai *et al.*, 2012), a forma da célula e a morfogênese da pétala (Oppenheimer *et al.*, 1991; Noda *et al.*, 1994), o desenvolvimento de tricomas (Wang *et al.*, 2010), a resposta a luz e deficiência de nutrientes (Stracke *et al.*, 2010; Dai *et al.*, 2012) e envolvidas no controle do metabolismo de fenilpropanóides, importante na defesa vegetal (Hichri *et al.*, 2011). TFs MYB desempenham papéis importantes na regulação de diversos metabólitos secundários, como as proantocianidinas na leguminosa modelo *Medicago truncatula* Gaertn. (Verdier *et al.*, 2012) e também já foram descritos como reguladores negativos em *Arabidopsis* (Jin *et al.*, 2000) e na alface (*Lactuca sativa* L.) (Park *et al.*, 2008).

Fatores MYB também participam na resposta das plantas a estresses ambientais e em resposta a ações hormonais, atuando em vias de transdução de sinais do ácido salicílico (AS) (Raffaele *et al.*, 2006), do ácido abscísico (ABA) (Abe *et al.*, 2003), do ácido giberélico (GA3) (Murray *et al.*, 2003) e de ácido jasmônico (JA) (Lee *et al.*, 2001). Sob déficit hídrico, o fitohormônio ABA induziu uma cascata de sinalização ativando os genes AtMYB60 e AtMYB96 em *Arabidopsis* para regular o movimento dos estômatos, sendo também verificada a

atuação destes genes na resistência a doenças (Seo *et al.*, 2009; Seo e Parque, 2010). Liao *et al.* (2008) identificaram 156 genes GmMYB em soja, dos quais 43 genes apresentaram expressão alterada no tratamento com ABA, sal, seca e/ou estresse causado pelo frio. Em *Arabidopsis*, genes MYB foram relacionados com a resposta de defesa da planta frente ao ataque de afídeos (Liu *et al.*, 2010) e insetos herbívoros (De Vos *et al.*, 2006).

2.2. A família Fabaceae e sua importância

A família Fabaceae, cujos membros são comumente conhecidos como leguminosas, figura entre as três famílias de plantas superiores com maior riqueza de espécies, abrigando cerca de 700 gêneros e aproximadamente 20.000 espécies. É um dos grupos mais diversificados da flora mundial, com distribuição cosmopolita e espécies nativas em todos os continentes, à exceção da Antártida (Lewis *et al.*, 2005). Com base nas diferenças morfológicas, as espécies da família Fabaceae distribuem-se em três subfamílias: a Caesalpinioideae, com aproximadamente 2.700 espécies de ambientes tropicais e subtropicais, sendo a maioria arbórea tropical; a Mimosoideae, abrigando em torno de 2.500 espécies geralmente arbóreas de regiões de climas tropicais, subtropicais e temperados; e a Papilionoideae (ou Faboideae), que representa o grupo mais numeroso com cerca 12.600 espécies em sua maioria herbáceas, mas também arbóreas, difundidas em regiões temperadas e ambientes tropicais (Judd *et al.*, 2002). Os representantes desta família botânica possuem hábitos de crescimento variados, existindo árvores de pequeno, médio e grande porte, ervas anuais ou perenes, arbustos e lianas, entre outros (Souza e Lorenzi, 2005; Souza e Souza, 2011).

Na flora brasileira, Fabaceae apresenta-se como a família mais rica em diversidade de espécies e figura entre as cinco com maior número de indivíduos do país. São registrados para o Brasil cerca de 211 gêneros e aproximadamente 2700 espécies de leguminosas distribuídas em todos os domínios fitogeográficos, estimando-se que o Cerrado seja o bioma que abriga maior diversidade de espécies já descritas (28,68%), seguidos pela Amazônia (27,57%), Mata Atlântica (23,25%), Caatinga (15,28%), Pampa (2,69%) e Pantanal (2,54%) (Lima *et al.*, 2012).

A importância dos processos ecológicos e biológicos das leguminosas é atestada por sua ampla distribuição em praticamente todos os tipos vegetacionais do mundo. No entanto, a riqueza deste grupo não se resume apenas ao grande número ou distribuição de suas espécies. O potencial econômico da família é bem acentuado, ficando atrás apenas de Poaceae, incluindo espécies exploradas pela medicina popular; na produção de óleo e resina, cortiça, lenha e carvão; como plantas ornamentais, forrageiras e madeireiras; pela indústria de alimentos e cosméticos; além de compreender espécies consideradas como modelos vegetais (*M. truncatula*, *G. max* e *Lotus japonicus* (Regel) K. Larsen), auxiliando no desenvolvimento científico (Lewis *et al.*, 2005; Souza e Souza, 2011).

Adicionalmente, muitas espécies de leguminosas são utilizadas para o melhoramento do solo, pois se associam simbioticamente com um grupo de bactérias fixadoras do nitrogênio, permitindo seu uso pelos agricultores na adubação verde (Souza e Souza, 2011). A família ainda tem grande destaque na dieta alimentar humana, perdendo em importância apenas para os cereais (Graham e Vance, 2003), pois representam excelentes fontes de proteínas e minerais essenciais, produtores de compostos secundários de promoção à saúde,

bem como de óleo vegetal processado para consumo humano, além de compreender um terço de todo o nitrogênio da proteína na dieta (Vance *et al.*, 2000; Grusak, 2002; Graham e Vance, 2003; Gepts *et al.*, 2005).

2.2.1. A cultura do feijão-caupi (*Vigna unguiculata*)

O feijão-caupi, popularmente conhecido como feijão-macassar, feijão verde ou feijão-de-corda, apresenta destaque na economia das regiões Norte e Nordeste do Brasil por constituir-se como um dos principais componentes da dieta da população, principalmente a de baixa renda, além de tratar-se de uma das principais fontes de emprego e renda dessas regiões (Lima *et al.*, 2007; Freire-Filho *et al.*, 2011). É uma cultura amplamente adaptada, versátil e nutritiva, constituindo-se como excelente fonte de proteínas, vitaminas, sais minerais e fibras (Timko *et al.*, 2007). Sua plasticidade, ampla variabilidade genética e boa adaptação a diversos ambientes, associados ao potencial produtivo dos grãos e boa capacidade de fixação de nitrogênio atmosférico, conferem à cultura um significativo valor estratégico, diminuindo assim a dependência do feijão comum, mais propenso a mudanças nos índices de produtividade devido a oscilações no regime pluviométrico (Freire-Filho *et al.*, 2011).

Atualmente, a cultura encontra-se entre as 23 espécies cultivadas de maior importância global, sendo o Brasil o terceiro maior produtor da cultura, a qual ocupa cerca de 11 milhões de hectares distribuídos pela Ásia, África, sul da Europa e América (Singh *et al.*, 2002). Segundo dados da CONAB (2012), a área cultivada com feijão-caupi no Brasil atinge cerca de 1,3 milhões de hectares com a produção de 480 mil toneladas. No Brasil, trata-se do único tipo de feijão capaz de se desenvolver tanto na região Norte, onde a umidade é alta e o solo

predominante é argiloso, como na Nordeste, onde o clima é mais seco e solo que prevalece é o arenoso. Admitindo-se que cada hectare de feijão-caupi gera 0,8 emprego/ano e considerando o consumo per capita médio do Brasil de 18,21 kg/pessoa/ano e o preço mínimo da saca de 60 kg de R\$ 80,00, constata-se que a cultura gerou, em média, 1.113.109 empregos por ano; produziu suprimento alimentar para 28.205.327 pessoas; e gerou uma produção anual no valor de R\$ 684.825.333,00 (Freire-Filho *et al.*, 2011).

Apesar da sua ampla variabilidade genética para praticamente todos os caracteres de interesse agrônômico, alguns acessos da cultura do feijão-caupi não apresentam resistência/tolerância aos principais estresses ambientais que atingem a cultura (Freire-Filho *et al.*, 2005). Entre tais estresses, o ataque de pragas e doenças, sobretudo aquelas causadas por vírus, podem acarretar em perdas substanciais da produção que, a depender da cultivar plantada, podem atingir até 80% da lavoura plantada. Dentre as principais viroses, o vírus do mosaico severo do caupi (*Cowpea severe mosaic virus* – CPSMV) se destaca pela severidade do seu ataque e o impacto sobre a produtividade da cultura (Lima *et al.*, 2005).

Atualmente os projetos de melhoramento da cultura visam ao desenvolvimento de cultivares com alta qualidade de grão, resistência à seca e à salinidade, resistência múltipla a viroses, doenças fúngicas e bacterianas, bem como porte mais compacto e ereto, que possibilitem a colheita mecânica e características para processamento industrial (Freire-Filho *et al.*, 2005). O melhoramento genético produziu progressos consideráveis nas espécies cultivadas no Brasil, estando entre os melhores países do mundo com

contribuições expressivas no que se refere à pesquisa com recursos genéticos e melhoramento genético vegetal (Freire-Filho *et al.*, 2011).

2.3. Genômica funcional do feijão-caupi: Banco de dados NordEST

Nos últimos anos, o feijão-caupi tem sido objeto de projetos que objetivam o sequenciamento do seu genoma e transcriptoma. Os projetos que trabalham no sequenciamento do genoma desta leguminosa têm disponibilizado suas sequências em bancos de dados públicos, como o CGKB (*Cowpea Genomics Knowledge Base*; Base de Conhecimentos Genômicos de Feijão-Caupi), um banco de dados baseados em informações derivadas de 298.848 sequências ricas em genes (*Cowpea Genespace Sequences - GSS*), geradas através da filtragem de DNA genômico metilado (Chen *et al.*, 2007); e o HarvEST, um banco de dados internacional, com mais de 180.000 ESTs geradas a partir de 17 bibliotecas oriundas de diversos tecidos (HarvEST, 2014).

Quanto ao sequenciamento do transcriptoma do feijão-caupi, em 2004 foi anunciado o projeto de genômica funcional, estrutural e comparativa do feijão-caupi – NordEST (rede brasileira de genômica do feijão-caupi), no âmbito do programa Renorbio/BNB (Benko-Iseppon *et al.*, 2005, 2008; Benko-Iseppon, 2009). Desde então, mais de 500 milhões de transcritos expressos foram gerados sob diferentes condições de estresses biótico (vírus do mosaico severo do caupi e vírus transmitido por afídeos) e abiótico (seca e salinidade), por meio de sequenciamento de ESTs (*Expressed Sequence Tags*) e tags SuperSAGE (*Super Serial Analysis of Gene Expression*), havendo também acesso a outras categorias

de sequências (cDNA *Full Length*, RNA-seq, miRNA, etc.), os quais revelaram genes-candidatos potencialmente úteis para o melhoramento da cultura.

Essa geração massiva de dados sobre a expressão de genes em diversas condições pode tornar a cultura do feijão-caupi altamente produtiva e rentável, sendo sua anotação um dos requisitos básicos para o entendimento da genômica funcional, ou seja, a chave para a interpretação biológica dos dados (Conesa *et al.*, 2005; Shivashankar *et al.*, 2006).

2.4. Análise global da expressão gênica

Os dados derivados do sequenciamento de genomas completos de diferentes organismos, embora indiscutivelmente relevantes, não esclarecem totalmente mecanismos moleculares envolvidos em muitos processos biológicos (Pandey e Mann, 2000). O acúmulo exponencial dessas sequências gênicas depositadas em bancos de dados públicos tem aumentado a demanda por metodologias que auxiliem na identificação funcional dos genes e/ou confirmação de homólogos, bem como na elucidação dos seus padrões de expressão (Calsa Junior *et al.*, 2004).

Mudanças na atividade transcricional de determinado gene são os principais indicativos de sua identidade e função, estando o seu padrão de expressão frequentemente correlacionado com a presença ou ausência de seus RNAs mensageiros (RNAm) correspondentes nas células. Genes que exibem padrões de expressão similares sob diferentes condições, possivelmente estão correlacionados funcionalmente, considerando-se que mudanças no perfil transcricional refletem demandas impostas sobre as células, alterando suas condições fisiológicas (Pollock, 2002). Sendo assim, as técnicas que permitem a

detecção dos níveis de mRNA na célula revelam o padrão de transcrição espaço-temporal dos genes, permitindo correlacioná-los com processos biológicos (Shimkets, 2004).

Diversas metodologias podem ser empregadas para quantificar a prevalência de mRNA de determinado gene em uma célula, as quais podem variar de forma significativa quanto ao tempo empregado para o desenvolvimento da técnica, os meios necessários para sua realização e, o mais importante, a sensibilidade do método adotado (Shimkets, 2004). Tais métodos podem ser divididos de forma geral em duas abordagens distintas, as quais se sobrepõem dependendo das necessidades do estudo em questão. Considerando-se o princípio metodológico empregado, as análises de expressão gênica diferenciam-se em três grupos: (I) aqueles baseados na hibridização de sondas; (II) por meio do sequenciamento de regiões específicas de fragmentos de cDNA e (III) através da análise de fragmentos de cDNA amplificados via PCR (Calsa Junior *et al.*, 2004).

Tais metodologias ainda podem ser divididas em outras duas categorias: (i) os métodos de análise abertos (globais), onde não é exigido o conhecimento prévio dos transcritos de interesse, permitindo desta forma o acesso a potencialmente todos os transcritos expressos em um determinado momento, e (ii) os métodos fechados, onde transcritos de interesse são já conhecidos, utilizando-se para tais análises sondas e/ou *primers* específicos. Os métodos por hibridização são considerados fechados, enquanto que os que analisam fragmento de cDNA amplificados ou sequenciados são caracteristicamente métodos abertos (Alba *et al.*, 2004; Calsa Junior *et al.*, 2004).

Entre as tecnologias mais utilizadas para análises de expressão gênica, a metodologia de microarranjos (do inglês *microarray*; Schena *et al.*, 1995) também conhecida como plataforma de chips de nucleotídeos, destaca-se dentre os métodos baseados em hibridação, enquanto que as técnicas de *Expressed Sequence Tag* (EST) (Adams *et al.*, 1991), *Serial Analysis of Gene Expression* (SAGE) (Velculescu *et al.*, 1995) e seus derivados *Long Serial Analysis of Gene Expression* (LongSAGE) (Saha *et al.*, 2002), a *Super Serial Analysis of Gene Expression* (SuperSAGE) (Matsumura *et al.*, 2003) e mais recentemente RNA-seq (Mortazavi *et al.*, 2008) destacam-se entre os principais métodos baseados em sequenciamento.

2.4.1. ESTs – *Expressed Sequence Tags*

Desde a sua introdução em Adams *et al.* (1991), a utilização do método de sequenciamento de Etiqueta de Sequência Expressa (EST) tem se mostrado uma ferramenta com diversas aplicabilidades, além do seu objetivo primário para análises de expressão gênica. Dentre essas, incluem-se a descoberta de novos genes, a complementação em anotações genômicas, a identificação da estrutura gênica e a determinação da viabilidade de transcritos alternativos, facilitando as análises proteômicas (Nagaraj *et al.*, 2006; De Paula *et al.*, 2011).

De forma resumida, a técnica consiste na construção de uma biblioteca de cDNA a partir de uma população de mRNA do tecido de interesse. Os cDNAs produzidos são então inseridos em vetores de clonagem e posteriormente sequenciados para a geração dos ESTs. Em linhas gerais, as sequências são produzidas em grande escala em um processo relativamente barato, de forma que os ESTs gerados são selecionados aleatoriamente da biblioteca de cDNA,

não sendo editados, tratando-se de sequências relativamente curtas (entre 200 e 800 pares de bases) cujo sequenciamento é feito em um único passo (Nagaraj *et al.*, 2006). Com o sequenciamento de milhares de clones, o número de transcritos de diferentes genes pode ser enumerado com o auxílio de ferramentas bioinformáticas, onde a frequência de cada transcrito irá representar o perfil de expressão deste na amostra biológica em estudo (Terauchi *et al.*, 2008).

Apesar de sua ampla aplicabilidade em estudos genômicos, a metodologia de EST apresenta diversas limitações que restringem seu uso em determinadas aplicações. Um dos principais entraves da técnica refere-se à representação global dos genes em uma dada biblioteca, assim como a qualidade das sequências finais produzidas. A amostragem do mRNA para a geração das bibliotecas de cDNA é condicionada ao estágio de desenvolvimento, condição ambiental imposta, bem como o tecido ou organismo analisado. Adicionalmente, o protocolo de obtenção das ESTs apresenta em torno de 4% de erro, resultante principalmente do sequenciamento em passo único, o que confere à sequência uma baixa qualidade (Lorkowski e Cullen, 2003).

2.4.2. SuperSAGE

Análise serial da expressão gênica (SAGE) (Velculescu *et al.*, 1995) é uma técnica de análise transcriptômica que fornece dados qualitativos e quantitativos da expressão de um gene, sendo altamente sensível na detecção dos transcritos, incluindo aqueles desconhecidos ou de expressão rara, mas limitada pelo seu curto tamanho de 13-15 pares de bases (pb) que dificulta a identificação do gene de origem (Matsumura *et al.*, 2003; Wang, 2007). Para contornar tais problemas,

foram desenvolvidas versões melhoradas da técnica, dentre as quais a SuperSAGE (Matsumura *et al.*, 2003) é uma das mais eficientes e modernas.

A metodologia de SuperSAGE baseia-se na síntese de cDNA a partir do mRNA utilizando-se como iniciador uma sequência oligo-d(T), seguido pela digestão com a enzima *NlaIII* e ligação de um adaptador ao final do cDNA clivado. Posteriormente, os fragmentos com os adaptadores são digeridos pela *EcoP15I* (enzima de restrição do tipo III que corta a 26-27 pb da extremidade 3' do local de reconhecimento), sendo purificados via eletroforese de gel de poliacrilamida para posterior sequenciamento (Matsumura *et al.*, 2008) (Figura 3).

A metodologia fundamenta-se no pressuposto de que cada tag gerada contém a informação suficiente para identificar apenas um gene, considerando-se sua obtenção de um único local do transcrito. Este aumento na extensão da tag gerada aumenta significativamente a eficiência da identificação dos genes correspondentes às tags (Matsumura *et al.*, 2008) possibilitando novas abordagens de estudos aplicando a técnica de SuperSAGE, dentre as quais a observação de um transcriptoma de interação, onde duas ou mais espécies estão interagindo (como por exemplo as interações de parasita e hospedeiro descritas por Matsumura *et al.*, 2003) ou para confecção de chips microarray que podem ser utilizados, assim como a RT-qPCR, para validação do perfil gerado pela SuperSAGE (Matsumura *et al.*, 2004; 2008).

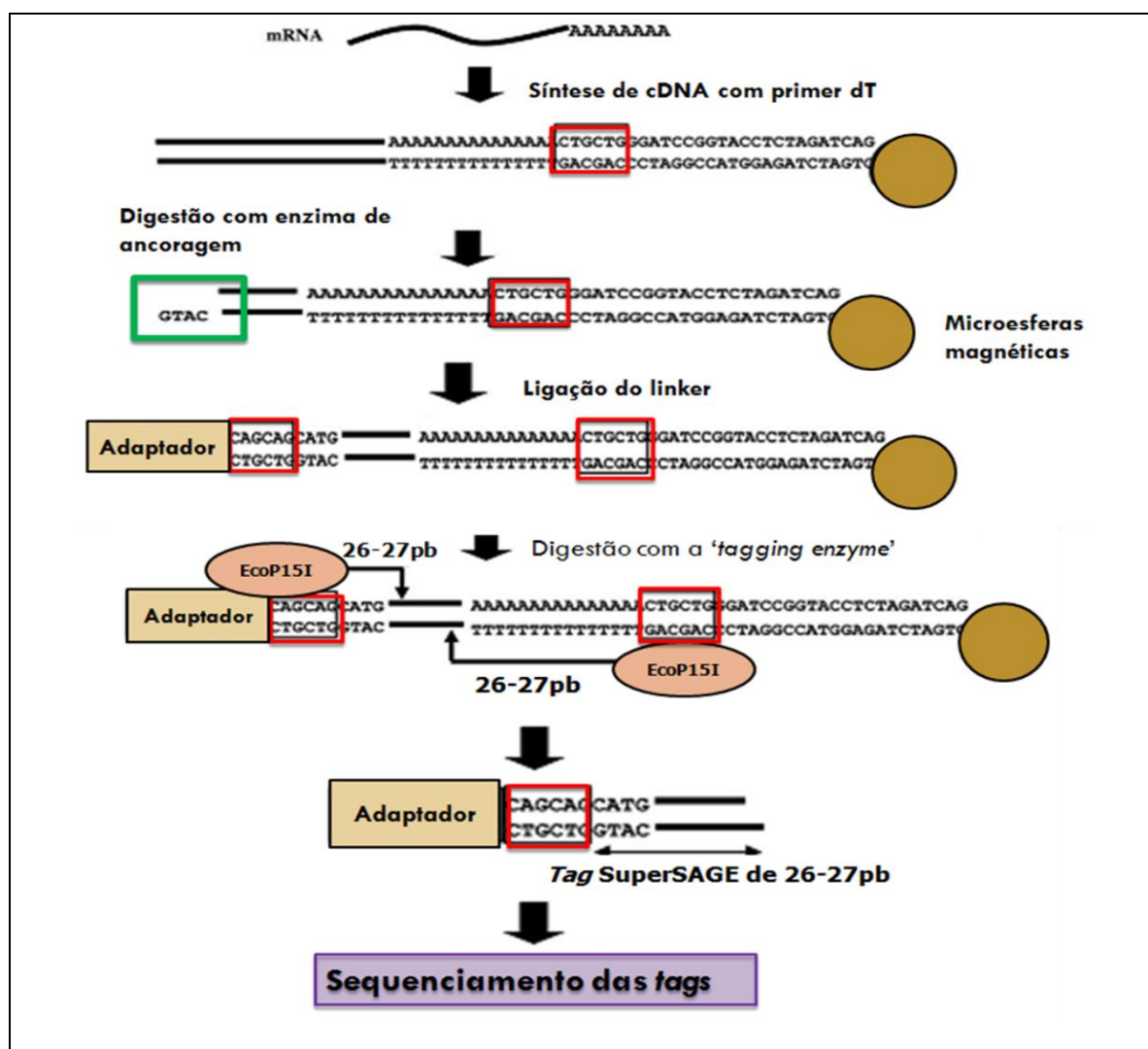


Figura 3: Fluxograma da técnica SuperSAGE. (Fonte: Traduzido de Matsumura *et al.*, 2008).

Outro fator relevante ao uso da metodologia de SuperSAGE refere-se à sua adaptação com as novas tecnologias de sequenciamento, originando atualizações da técnica com maior rendimento e eficiência nas análises, como a HT-SuperSAGE (Matsumura *et al.*, 2010) e DeepSuperSAGE (Matsumura *et al.*, 2012). Esta adequação da SuperSAGE resultou numa diminuição de custo e de tempo de produção ao proporcionar uma ampliação na amostragem das tags do transcriptoma analisado.

2.4.3. RNA-seq

Trata-se de uma técnica relativamente recente e em rápido desenvolvimento denominada RNA-seq (Sequenciamento do RNA ou também chamada *Whole Transcriptome Shotgun Sequencing* - WTSS) a qual se vale de tecnologias de sequenciamento de alto rendimento para determinar um nível mais exato de transcritos expressos em uma amostra (Mortazavi *et al.*, 2008; Wolf, 2013). Essa recente metodologia apresenta diversas vantagens sobre as tecnologias existentes, como exemplo, o RNA-seq não se limita ao conhecimento prévio do genoma ou dos transcritos de um organismo, mostrando-se como uma tecnologia bem atrativa para espécies cuja sequência genômica não foi previamente determinada (Wang *et al.*, 2009b). Ademais, a tecnologia promete desvendar complexidades anteriormente inacessíveis ao transcriptoma, tais como a identificação de expressão alelo-específica, isoformas de genes, identificação de mutações, novos promotores e a localização exata dos limites da transcrição (Wang *et al.*, 2009b; Oshlack *et al.*, 2010).

De uma forma simplificada, a técnica se baseia na conversão de uma população de RNA (total ou fracionado, como poli (A)+) em uma biblioteca de fragmentos de cDNA. Posteriormente, são ligados adaptadores a uma ou ambas as extremidades destes fragmentos, levados para sequenciamento de alto rendimento, com ou sem amplificação prévia, gerando ao final sequências (*reads*) curtas na ordem de 30 a 400 pb. Após o sequenciamento, as *reads* resultantes são alinhadas a um genoma ou transcriptoma de referência, ou até mesmo montadas de novo, a fim de produzir um mapa transcriptômico de escala genômica composto pela estrutura transcricional e/ou nível de expressão de cada gene (Figura 4) (Wang *et al.*, 2009b).

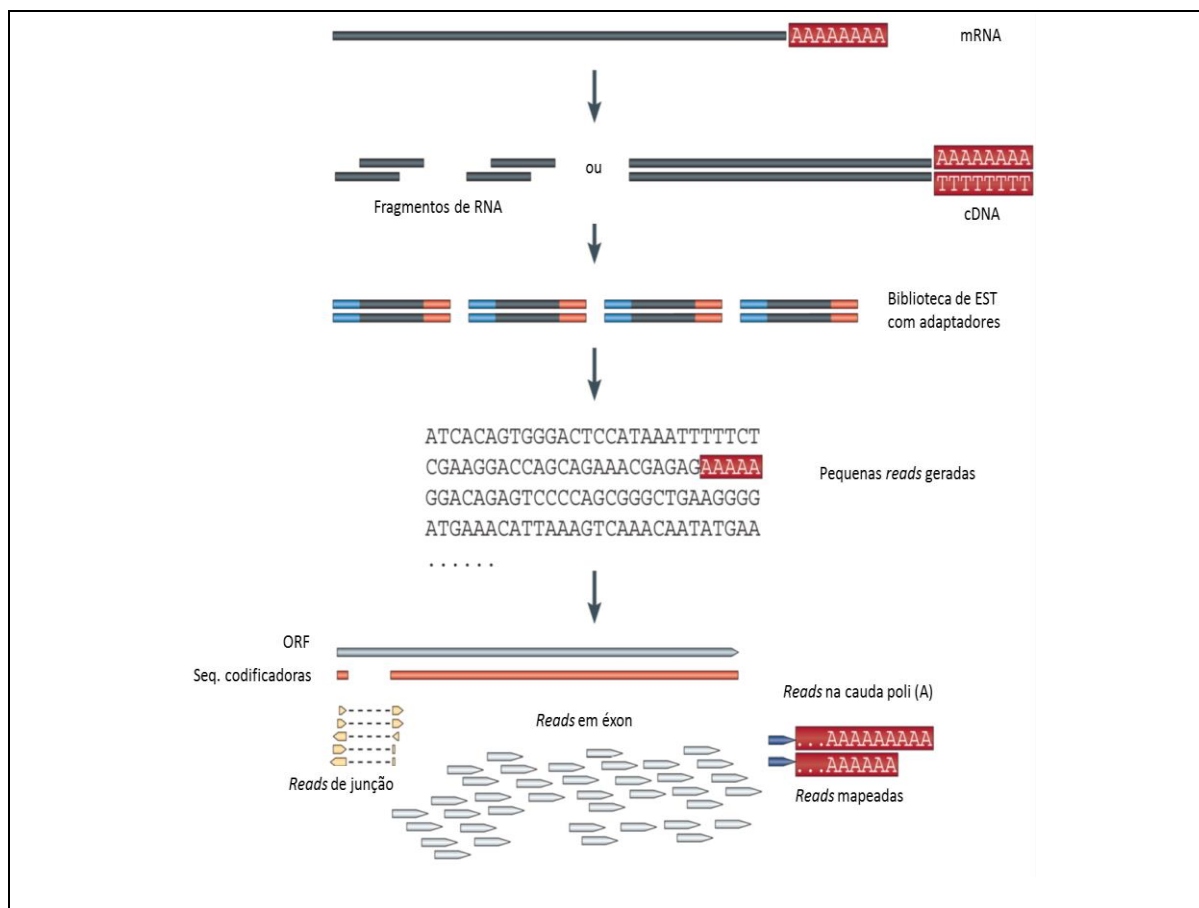


Figura 4: Construção de uma biblioteca de cDNA. Os RNAs longos primeiramente são convertidos em uma biblioteca de fragmentos de cDNA (por meio da fragmentação do RNA ou do DNA). Subsequentemente, os adaptadores de sequenciamento (azul) são adicionados a cada fragmento de cDNA e uma sequência curta é obtida utilizando tecnologia de sequenciamento de alto rendimento. As *reads* resultantes são alinhadas com o genoma ou transcriptoma de referência e classificadas em três tipos: *reads* em éxon, *reads* de junção e *reads* na cauda poli (A). (Fonte: Traduzido de Wang *et al.*, 2009b).

A tecnologia RNA-seq tem se consolidado como um método padrão para análise transcriptômica, mas ainda possui inúmeros desafios que permeiam o desenvolvimento da técnica como, por exemplo, a necessidade de um *pipeline* de bioinformática altamente eficiente para armazenar, acessar e principalmente processar o imenso conjunto de dados gerados (Palmieri e Scotterer, 2009). De toda forma, o RNA-seq é o primeiro método baseado em sequenciamento que

permite que todo o transcriptoma possa ser estudado com um alto rendimento e de forma quantitativa (Wang *et al.*, 2009b).

2.4.4. RT-qPCR

Apesar da disponibilidade de uma variedade de métodos para quantificar a expressão de um gene, a PCR quantitativa em tempo real (RT-qPCR) é considerada atualmente como o “padrão-ouro” para analisar a quantidade exata de um único transcrito de mRNA numa amostra biológica (Sinicropi *et al.* 2007; Vanhauwaert *et al.*, 2014). A alta sensibilidade, especificidade, exatidão e precisão da técnica, bem como sua simplicidade e velocidade de processamento prático, tornam a RT-qPCR o melhor método de escolha para a validação de estudos genômicos (Vanhauwaert *et al.*, 2014).

A técnica consiste na amplificação de um produto (ácido nucléico) e detecção dos dados durante a reação, ou seja, em tempo real, por meio da captação de um sinal fluorescente específico (sondas marcadas) ou não específico (intercalantes de DNA). O Sybr Green é um exemplo de fluoróforo de detecção não específica, pois se intercala a qualquer fita-dupla de DNA presente na amostra e emite fluorescência. Por sua vez, o TaqMan é um exemplo de sonda de detecção específica que distingue as sequências alvo de interesse dos produtos não específicos amplificados. Esse tipo de sonda permite ainda a quantificação de múltiplos genes em uma mesma reação, reduzindo dessa forma a quantidade de amostras e reagentes necessários. Apesar de todas essas vantagens, o Sybr Green ainda é o mais intensivamente utilizado pelos pesquisadores, principalmente devido ao seu baixo custo e facilidade no desenvolvimento do ensaio (Bio-Rad Laboratories, 2006; Shipley, 2006).

Apesar da robustez do método, algumas variáveis existentes durante o processo podem afetar os resultados da RT-qPCR, tais como a extração e armazenamento do RNA, a transcrição reversa, a especificidade dos *primers* na reação, a eficiência de cada curva de amplificação, entre outros (Lander *et al.*, 2001; Venter *et al.*, 2001). Para corrigir tais variações existentes é importante aplicar uma boa estratégia de normalização. Um dos procedimentos mais frequentemente aplicados à técnica é o uso de genes normalizadores internos adequados para cada condição avaliada (Bustin *et al.*, 2009). Os genes de referência devem mostrar níveis de expressão invariáveis nas amostras estudadas apresentando, portanto, uma forte correlação com a quantidade total de mRNA presente nas amostras. Com isso, pesquisas que utilizam tal tecnologia, agregam mais uniformidade e confiabilidade aos dados obtidos (Bustin *et al.*, 2009).

3. Objetivos

3.1 Geral

- ✓ Identificar, caracterizar e classificar *in silico* os TFs da família MYB do feijão-caupi e validar os padrões de expressão dos melhores candidatos envolvidos nos mecanismos de resposta à defesa contra infecção pelo vírus do mosaico severo do caupi (*Cowpea severe mosaic virus* – CPSMV), bem como estabelecer genes de referência para estudos de expressão diferencial sob diferentes condições de estresse.

3.2 Específicos

- ✓ Selecionar TFs da família MYB relacionados com estresses bióticos na literatura em outras leguminosas para identificação de ortólogos.
- ✓ Identificar e caracterizar *in silico* os genes codificantes dessa família de TFs no banco de dados do feijão-caupi - NordEST.
- ✓ Analisar a evolução de genes selecionados comparativamente com suas sequências seed (*Arabidopsis*).
- ✓ Selecionar tags SuperSAGE diferencialmente expressas à infecção pelo CPSMV e validá-las via RT-qPCR.
- ✓ Determinar genes de referência para estudos de expressão diferencial frente ao estresse hídrico, salino e viral.

4. Capítulo I

Artigo a ser submetido à revista *Genetics and Molecular Biology*

**Expressão Diferencial e Diversidade de Fatores de Transcrição
da Família MYB no Transcriptoma do Feijão-Caupi
[*Vigna unguiculata* (L.) Walp.]**

Mitalle K. S. Matos¹; Flávia T. Araújo¹; Lidiane L. B. Amorim¹; João P. Bezerra-Neto¹; Ana M. Benko-Iseppon¹

¹Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego 1235, CEP 50.670-420, Recife, PE, Brasil.

E-mail do autor para correspondência: ana.iseppon@gmail.com

RESUMO

O controle sobre a expressão gênica é responsável pela regulação de praticamente todos os processos biológicos nos vegetais, incluindo os mecanismos de resposta a diferentes estímulos provindos do ambiente, entre estes, os estresses causados por infecções virais. Sob condições de estresse, os fatores de transcrição (TFs) participam ativamente das etapas iniciais do processo de percepção e sinalização, podendo atuar como ativadores ou repressores da expressão gênica. A doença provocada pelo mosaico severo do caupi (CPSMV) é responsável por perdas significativas na produção de uma das leguminosas de maior importância socioeconômica do Norte e Nordeste brasileiro, o feijão-caupi. Neste contexto, o presente trabalho objetivou caracterizar *in silico* TFs da família MYB no transcriptoma do feijão-caupi e analisar a regulação de sua expressão frente ao ataque do CPSMV. Por meio de análises *in silico* utilizando dados transcriptômicos obtidos no banco NordEST, identificamos um total de 86 candidatos a TFs MYB, classificados nas subfamílias R1-MYB (20), R2R3-MYB (63) e R3-MYB (3). A análise dos componentes estruturais do domínio R2R3-MYB permitiu observar a conservação dos aminoácidos característicos desta classe proteica dentro da espécie. O padrão de distribuição genômico foi determinado pela ancoragem em pseudocromossomos de *Phaseolus vulgaris* L., apoiando fortemente a hipótese de que processos de duplicação segmentar e em *tandem* contribuíram para a expansão dos TFs MYB. A análise fenética utilizando as sequências do domínio R2R3-MYB formou 18 subclados, os quais foram apoiados pelos motivos conservados da região C-terminal, refletindo a conservação funcional dos grupos formados. Das tags SuperSAGE diferencialmente expressas sob a infecção pelo CPSMV, três tiveram sua

expressão induzida com o decorrer do tempo de estresse (16 h), indicando a participação ativa dessa família na resposta de defesa. Por outro lado, um dos genes analisados teve sua expressão negativamente regulada nos tempos iniciais do estresse. Os resultados aqui obtidos são pioneiros para a espécie estudada e mostram-se como excelente fonte de dados para estudos posteriores de validação e melhoramento da espécie.

Palavras-chave: Bioinformática; Estresse biótico; Defesa vegetal; Perfil transcricional; Leguminosa.

INTRODUÇÃO

A família MYB de fator de transcrição (*Transcription Factor* – TF) constitui-se como uma das classes mais abundantes de TFs específicos dos vegetais, estando presente em uma vasta gama de organismos, incluindo plantas terrestres (Hou *et al.*, 2014). Caracterizada pela presença de um domínio de ligação ao DNA altamente conservado, as proteínas MYB geralmente são compostas de uma a quatro repetições imperfeitas. Resíduos de Triptofano espaçados regularmente entre si são característicos de uma repetição MYB, onde cada repetição possui cerca de 50 aminoácidos (aa) codificando três α -hélices que se conformam em uma estrutura terciária em hélice-volta-hélice (*helix–turn–helix* – HTH) responsável pela ligação ao DNA (Stracke *et al.*, 2001; Du *et al.*, 2009; 2012a). Com base na quantidade de repetições imperfeitas do domínio MYB, diferentes subfamílias podem ser identificadas, sendo estas: R1-MYB, R2R3-MYB, R3-MYB e R4-MYB, contendo um, dois, três ou quatro domínios, respectivamente (Stracke *et al.*, 2001; Dubos *et al.*, 2010). Genes R2R3-MYB constituem-se como a maior subfamília de TFs MYB em vegetais desempenhando diversas funções em processos específicos da planta (Stracke *et al.*, 2001; Du *et al.*, 2012a).

Funcionalmente diversificadas, as proteínas MYB já foram descritas atuando em inúmeros processos biológicos, tais como o desenvolvimento da planta, o metabolismo secundário, a transdução de sinais hormonais, resistência e tolerância aos diversos estresses ambientais, entre outros (Stracke *et al.*, 2001; Dubos *et al.*, 2010; Cai *et al.*, 2012). Inúmeros estudos em organismos vegetais têm sido conduzidos demonstrando o papel dos TFs MYBs nos mecanismos de resposta a estresses ambientais e na defesa a fitopatógenos (Ambawat *et al.*, 2013; Xie *et al.*, 2014). A maioria dos dados disponíveis provém de estudos em

Arabidopsis, onde 86 membros já tiveram suas potenciais funções estabelecidas, dos quais 18,25% atuam em respostas de defesa (Aoyagi *et al.*, 2014). Por exemplo, AtMYB96 e AtMYB44 foram induzidos em resposta a *Pseudomonas syringae* e promoveram, principalmente, a expressão de genes PR (*Pathogen Related*) (Seo e Park, 2010; Zhou *et al.*, 2013), ao passo que AtMYB31 foi identificado como sendo responsivo ao elicitor quitina (componente estrutural da parede celular de fungos) simultaneamente a outros 11 AtMYBs e genes de defesa (Libault *et al.*, 2007).

O feijão-caupi [*Vigna unguiculata* (L.) Walp] destaca-se por ser uma leguminosa de elevada importância socioeconômica principalmente para as regiões Norte e Nordeste do Brasil, tendo sua produtividade severamente afetada por doenças, sobretudo aquelas causadas por vírus, com as perdas podendo chegar a até 80% da lavoura plantada, a depender da cultivar e da época do ano (Booker *et al.*, 2005). Dentre os principais vírus que afetam a cultura, o vírus do mosaico severo do caupi (*Cowpea severe mosaic virus* – CPSMV) se destaca por sua severidade e impacto na produção. A Rede Nordestina de Biotecnologia (RENORBIO) realizou o estudo da funcional, estrutural e comparativa do feijão-caupi – NordEST (rede brasileira de genômica do feijão-caupi) sob diferentes condições de estresses biótico e abiótico e tem identificado potenciais candidatos importantes para o melhoramento do feijão-caupi, incluindo os TFs e genes reguladores de cascatas de sinais. Portanto, a identificação e estudo funcional de TFs MYB que regulam as respostas ao estresse viral no feijão-caupi podem elucidar os mecanismos moleculares por trás das respostas ao estresse, indicando genes candidatos potenciais para o melhoramento da cultura.

MATERIAL E MÉTODOS

Identificação da família MYB no feijão-caupi

Para a prospecção dos genes MYB no transcriptoma do feijão-caupi foram utilizadas 24 sondas (*seed sequence* – Anexo 1) da planta modelo *Arabidopsis thaliana* obtidas a partir do banco de dados *Plant Transcription Factor Database* (Zhang *et al.*, 2011). As proteínas foram previamente selecionadas com base na sua atuação em processos biológicos, de modo que fossem responsivas a agentes patogênicos e/ou moléculas sinalizadoras em estresses bióticos (ácido salicílico, ácido jasmônico, entre outros).

Uma busca via tBLASTn foi realizada na base de dados do NordEST/UFPE (<http://bioinfo03.ibi.unicamp.br/vigna/>). Para a triagem das ESTs (*Expressed Sequence Tag*) utilizou-se um ponto de corte (*e-value*) menor ou igual que e^{-5} . Os candidatos obtidos foram submetidos a uma clusterização pelo software CodonCode Aligner (<http://www.codoncode.com/index.htm>) para remover possíveis redundâncias entre as sequências. Os novos *contigs* (sequência consenso) e *singlets* gerados foram submetidos a alinhamentos recíprocos e reversos contra o banco de sequências não redundantes do NCBI (*National Center for Biotechnology Information*) através do algoritmo BLASTx (Altschul *et al.*, 1997), procedendo-se com a caracterização de suas ORFs (*Open Reading Frames*), utilizando o programa ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf>), e tendo os domínios conservados determinados por meio do algoritmo CD-search/rps-BLAST. Foram excluídas da análise sequências com as ORFs e/ou domínios incompletos, bem como aquelas que não puderam ser alinhadas por conterem regiões extensas de aa não definidos.

Alinhamento múltiplo das sequências R2R3-MYB

Para analisar as características das proteínas R2R3-MYB identificadas no transcriptoma do feijão-caupi, foi realizada uma análise comparativa dos seus domínios MYB através da geração de um alinhamento múltiplo utilizando o Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), sendo visualizados pelo Jalview v.2.8 (<http://www.jalview.org/>). A fim de obter o alinhamento otimizado, foram excluídos desta análise sete candidatos que apresentaram um padrão de espaçamento distinto entre as repetições R2 e R3 da proteína.

Mapeamento Circos

As 86 sequências candidatas a TF MYB do feijão caupi foram ancoradas em pseudocromossomos de *Phaseolus vulgaris* via BLASTn no banco de dados Phytozome, com o objetivo de inferir sobre sua distribuição nos cromossomos virtuais. Em seguida, as posições de ancoragem identificadas foram submetidas ao programa de Circos (Krzywinski *et al.*, 2009), sendo assim editado para gerar uma imagem de maior resolução das duplicações num gráfico baseado em uma organização circular dos pseudocromossomos do feijão-comum (n = 11).

Análise Fenética

Para análise fenética, as sequências de aa do domínio R2R3-MYB do feijão-caupi foram alinhadas com as sequências iscas de *A. thaliana* da mesma subfamília, com base no algoritmo MUSCLE disponível no programa MEGA 6 (Tamura *et al.*, 2013). O método de Neighbor-Joining foi empregado para construção da árvore, realizando uma análise de *bootstrap* com 1000 reamostragens.

Identificação de motivos fora do domínio MYB

Para a identificação do motivo conservado na porção C-terminal das proteínas MYB do feijão-caupi foi utilizada a ferramenta MEME (versão 4.9.1) (<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>) (Bailey *et al.*, 2006). As sequências de entrada compreenderam as regiões C-terminais que flanqueiam os domínios de proteínas R2R3-MYB do feijão-caupi e *Arabidopsis*. Os parâmetros definidos foram ajustados para: distribuição dos motivos, 0 ou 1 por sequência; largura mínima dos motivos, 6; largura máxima de motivo, 250; número máximo de motivos de encontrar, 50. Apenas motivos com um valor de e-value $\leq 1e^{-10}$ foram considerados.

Seleção das Tags SuperSAGE e Desenho dos *primers*

As tags disponíveis para o trabalho foram obtidas em bibliotecas SuperSAGE oriundas de folhas do genótipo BR-14 Mulato do feijão-caupi tratadas [30', 60', 90' (bulcados) e 16h] e não tratadas (controle), com o inóculo do CPSMV. As bibliotecas SuperSAGE foram geradas conforme Molina *et al.* (2008) e sequenciadas quase a totalidade via Solexa por intermédio da empresa GenXPro GmHb (Frankfurt, Alemanha). As tags foram extraídas e as unitags (tags diferentes) foram determinadas após terem sido excluídas aquelas sequenciadas somente uma vez (*singletons*), com a ajuda do software DiscoverySpace 4.0 (Robertson *et al.*, 2007).

As tags SuperSAGE diferencialmente expressas ao ataque do CPSMV foram selecionadas de duas formas: (1) por palavra-chave (MYB) das tags já anotadas e; (2) por meio de um BLASTn em um banco de dados local das tags supracitadas, utilizando como *seed sequence* os candidatos a proteínas MYB do

feijão-caupi. Foram consideradas para análise somente tags com alinhamentos perfeitos (tag – VuMYB) ou com no máximo um erro em 26 pares de bases (pb), não podendo este envolver nenhuma das quatro bases iniciais que caracteriza a tag (CATG). Ainda foram desconsideradas as tags que alinham com a mesma sequência, bem como aquelas que alinham em mais de uma (considerando o critério 1 tag = 1 VuMYB).

Dez tags foram selecionadas com base no seu nível de indução (*fold-change*) para o desenho dos *primers*, usando como *input* para o programa as sequências ESTs do banco NordEST que melhor alinham com as mesmas. A ferramenta utilizada foi o *Primer3* (<http://frodo.wi.mit.edu/primer3/>) seguindo os parâmetros pré-determinados pelo programa, exceto quanto ao seu conteúdo de GC (45 a 55%, com ótimo de 50%), tamanho do *amplicon* (entre 80 a 190 pb), temperatura de dissociação (TM) do produto (75 a 95 °C, com ótimo de 85 °C), máximo de auto complementaridade (4) e máximo de poli-X (3).

Material Biológico e Aplicação do estresse

Sementes do feijão-caupi (cultivar BR-14 Mulato) tolerante ao ataque do CPSMV foram concedidas pela Embrapa - Meio Norte (Teresina, PI) para o bioensaio de estresse viral. O experimento foi conduzido em fotoperíodo natural e temperatura variando entre 28 e 32° C, em telado antiafídeos na casa-de-vegetação do Instituto Agrônomo de Pernambuco (IPA). Após 21 dias da semeadura as plantas foram submetidas à inoculação mecânica com isolados do CPSMV procedendo-se com a coleta do material vegetal (folhas) após 30, 60 minutos e 16 horas da inoculação, sendo imediatamente congelado em N₂ líquido e armazenado a -80°C.

Extração do RNA total e síntese do cDNA

Aproximadamente 200 mg de material vegetal foi utilizada para a extração de RNA total através do kit “SV Total RNA Isolation System®” (Promega) de acordo com recomendações do fabricante. Após tratamento com DNase, 1 µg do RNA, quantificado via NanoDrop (Thermo Scientific, USA), foi utilizado para a síntese de cDNA utilizando o kit “QuantiTect® Reverse Transcription Kit” (Qiagen).

Reações de RT-qPCR

As reações de RT-qPCR foram realizadas utilizando o Kit "GoTaq® qPCR Master Mix" (Promega), seguindo as recomendações do fabricante mais otimização. A amplificação foi feita no equipamento Fluorescent Quantitative Quantification System (BIOER®) considerando os parâmetros utilizados por Daniel-Silva *et al.* (2013). A quantificação relativa, utilizando como genes de referência o ACT (actina), Beta-Tubulina e o SKIP, foi analisada pelo o software REST (Relative Expression Software Tool) versão 2.0.13 (Pfaffl *et al.*, 2002).

RESULTADOS

Identificação dos genes MYB

A busca por ortólogos no transcriptoma do feijão-caupi utilizando como sondas sequências protéicas de *Arabidopsis* retornou um total de 654 sequências candidatas contendo repetições MYB. Para eliminar possíveis redundâncias entre as sequências candidatas, as mesmas foram submetidas a uma reclusterização retornando 136 *contigs* e 62 *singletons* (sequências não alinhadas). Tais sequências foram provisoriamente nomeadas de acordo com os *contigs*

reclusterizados, recebendo o prefixo VuMYB. Foram excluídos da análise posterior, sequências com domínios e/ou ORFs incompletos na região N-terminal, bem como aquelas que não puderam ser alinhadas, por conterem regiões extensas de aa não definidos.

Por fim, obtiveram-se um total 86 candidatos a TFs MYB de feijão-caupi, sendo 74 *contigs* e 12 *singletons* confirmados como TFs MYB (Anexo 2). Tais sequências foram classificadas em subfamílias de acordo com o número de repetições do domínio MYB presentes, onde 20 candidatos apresentaram apenas um domínio conservado (1R MYB), 63 possuíam duas repetições MYB conservadas (R2R3-MYB) e três apresentaram três domínios MYB conservados (R3-MYB). A análise revelou que a subfamília R2R3 concentra o maior número de candidatos a genes desta família. Dada a grande importância da subfamília e sua predominância nas plantas, algumas análises no presente estudos foram centradas nesta subclasse de TFs.

Resíduos conservados no domínio MYB

Para analisar os componentes estruturais dos domínios homólogos e a frequência dos aa mais prevalentes em cada posição dentro de cada repetição do domínio R2R3-MYB do feijão-caupi, foi realizado um alinhamento com 56 proteínas da subfamília R2R3-MYB desta leguminosa (Figura 1). Em geral, a região básica do domínio das proteínas continha aproximadamente 110 resíduos de aa (incluindo a região de ligação) com deleções ou inserções raras (FF5457 / SI6968 / SI4072 / SI1579 / VuMYB_29 / 52 / 109 / 40). Em contrapartida, a região fora do domínio de ligação ao DNA foi divergente em termos de comprimento e composição de aa (dados não mostrados). As repetições R2 e R3 do feijão-caupi

continham aa característicos do domínio MYB, incluindo os resíduos de Triptofano (Trp – W) distribuídos uniformemente e conservados em posições específicas em ambas as repetições do domínio MYB. A repetição R3 mostrou-se mais conservada em número e composição de aa, quando comparada à R2, na qual foi observado que 14% das proteínas alinhadas continham inserções ou deleções de aa nas posições 21, 25, 35, 36 e 37 da repetição R2.

Com exceção das proteínas com indels em seu domínio R2, as repetições MYB do feijão-caupi apresentaram os seis resíduos de W característicos do domínio, espaçados regularmente nas posições 6, 28 e 50 da repetição R2, e 63, 82 e 101 da repetição R3, formando um centro hidrofóbico. No geral, à ressalva do Trp-63, o nível de conservação dos resíduos de W nos domínios MYB é alto, com apenas três proteínas (VuMYB_SI68696, VuMYB_1 e VuMYB_61) apresentando substituição no terceiro W da repetição R3. Na posição Trp-63, 95% das proteínas traziam uma substituição por outro aa, a saber, Phe (F), Ile (I) ou Leu (L).

Além dos resíduos Trp altamente conservados, observou-se a presença de aa alternativos também conservados em mais de 90% nas posições das repetições MYB do feijão-caupi. Estes incluíram Gli-4, Glu-10, Asp-11, Leu-14, Gli-23, Arg-41, Lys-44, Ser-45, Cys-46, Arg-47, Leu-48, Arg-49, Asn-52 e Leu-54 na repetição R2, e Pro-56, Glu-67, Gly-79, Ile-85, Ala-86, Pro-90, Gly-91, Arg-92, Thr-93, Asp-94, Asn-95, Lys-98, e Asn-99 na repetição R3 (Figura 1).

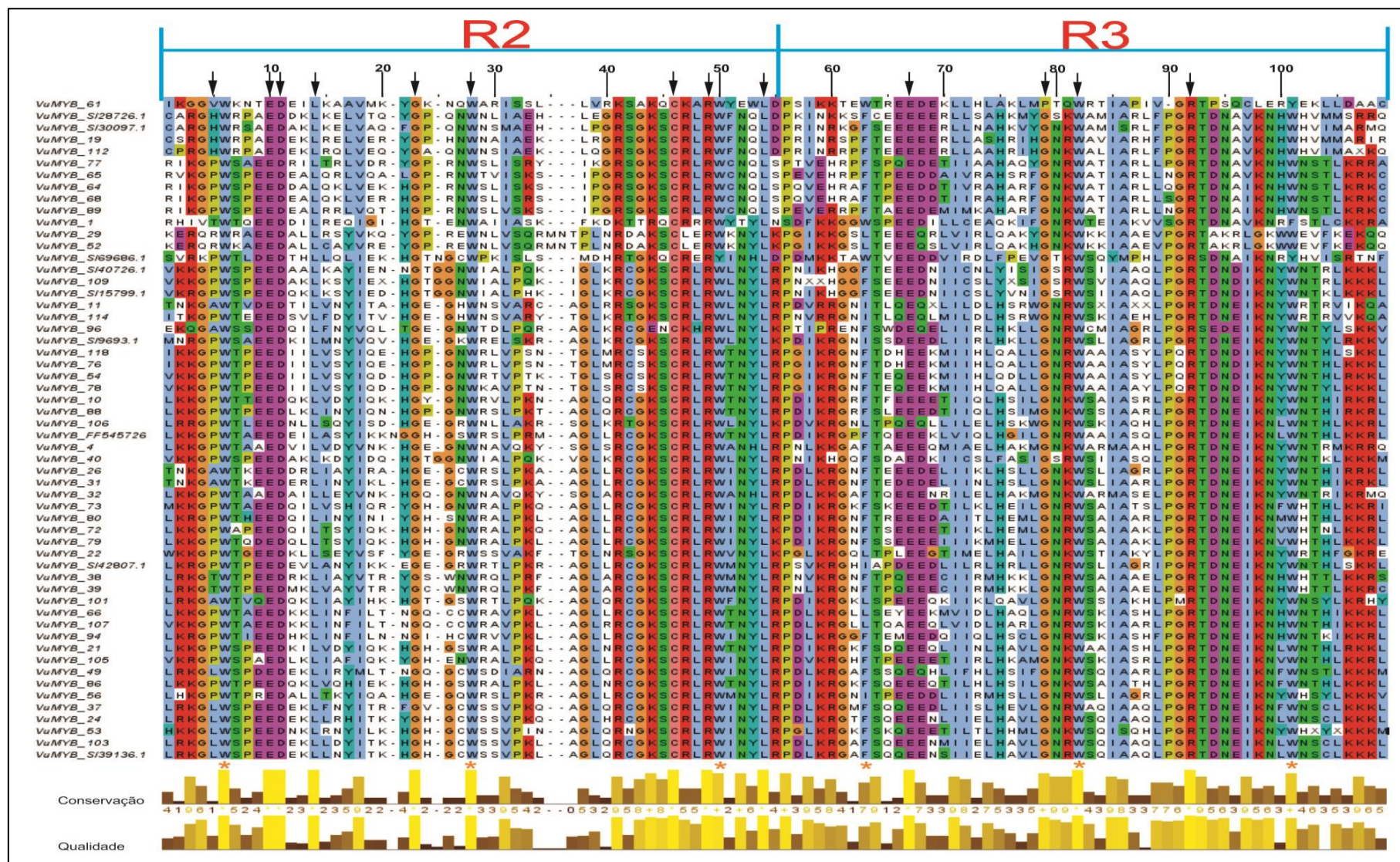


Figura 1: Alinhamento da sequência de aminoácidos de 59 genes com domínio R2R3-MYB do feijão-caupi utilizando o Clustal Omega. As setas indicam os aminoácidos conservados em todos os genes e os asteriscos indicam os resíduos de Triptofano que formam o centro hidrofóbico da proteína. Gráficos na base da figura indicam a conservação e a qualidade do alinhamento.

Distribuição cromossômica e eventos de duplicação em proteínas VuMYB

Para inferir sobre a distribuição genômica dos TFs MYB do feijão-caupi, a posição dos 86 candidatos foi mapeada em pseudocromossomos de *P. vulgaris*. A ancoragem destas sequências revelou que essa família de TF distribui-se por todos os 11 cromossomos virtuais do feijão-comum, mapeando-se um total de 163 locos (Figura 2). A proteína FF5457 não pôde ser mapeada de forma conclusiva em qualquer cromossomo. No geral, foi possível observar que a densidade destas proteínas não foi uniforme nos cromossomos analisados, variando de 4 a 18%. O cromossomo 03 reportou o maior número de genes ancorados, com 30 locos, seguido pelo cromossomo 08, com 20 genes correspondentes. Nos cromossomos 06 e 10 foram ancorados 07 locos correspondentes aos TFs MYB, representando os cromossomos com menor número de genes ancorados.

Densidades relativamente altas de MYBs foram observadas em algumas regiões cromossômicas, incluindo a parte superior e inferior dos cromossomos 02, 03, 07 e 08; o topo do cromossomo 11; e na porção inferior dos cromossomos 01, 06 e 09 para a subfamília R2R3 (Figura 2a). Para as subfamílias R1 e R3-MYB, apenas o cromossomo 03 mostrou-se com maior densidade de TFs MYB ancorados em ambas as regiões do cromossomo, enquanto que os cromossomos 04 e 09 foram mais densos no braço maior e apenas o 08 apresentou uma densidade de genes alta na região do braço menor (Figura 2b). Menos frequentemente, regiões pericentroméricas foram ricas em genes (por exemplo, nos cromossomos 2, 3 e 11) ou regiões intercalares.

Figura 2: Localizações cromossômicas e regiões de duplicação das proteínas MYB do feijão-caupi. O número do cromossomo é indicado na região externa e a posição dos seus centrômeros por uma linha branca. Linhas coloridas indicam supostas duplicações das sequências MYB; **(A)** Distribuição e duplicações das proteínas pertencentes às subclasses R2R3-MYB e **(B)** R1-MYB e R3-MYB.

No presente estudo, a duplicação intracromossômica também foi observada entre os genes MYB do feijão-caupi. Algumas duplicações em tandem de genes estreitamente relacionados em um único agrupamento foram observadas nos cromossomos 07 (VuMYB_SI96), 04 (VuMYB_SI30), 06 (VuMYB_32) e 10 (VuMYB_29) para a subfamília R2R3-MYB e nos cromossomos 09 (VuMYB_30), 04 (VuMYB_110) e 08 (VuMYB_91) para a subfamília R1-MYB.

Caracterização Fenética e Distribuição em Subgrupos

Para avaliar as relações entre as proteínas R2R3-MYB do feijão-caupi e *Arabidopsis* foi gerada uma árvore fenética utilizando o método de NJ. Com base na similaridade das sequências e topologia da árvore, os agrupamentos formados foram divididos em 18 subgrupos designados de C1 a C18, de acordo com os clados que apresentaram, pelo menos, 50% de apoio de *bootstrap* (Figura 3a).

Apesar dos baixos valores de suporte para nós internos da árvore, os nós mais externos receberam valores de *bootstrap* mais altos. Para a classificação dos subgrupos das proteínas R2R3-MYB, também foram considerados os resultados obtidos por Stracke *et al.* (2001) e Dubos *et al.* (2010) para os AtMYBs utilizadas como sondas no presente estudo.

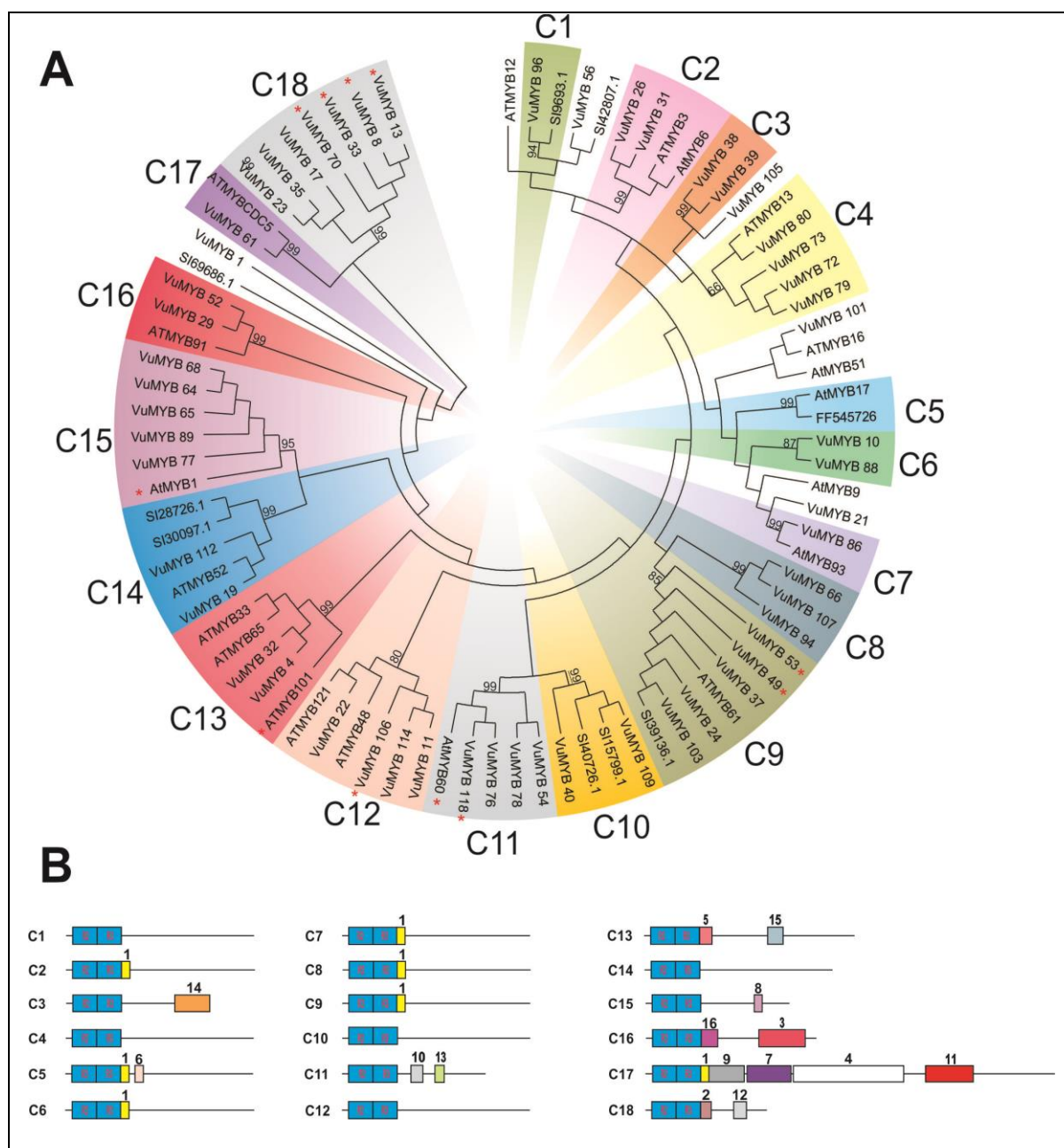


Figura 3: Dendrograma refletindo padrões da estrutura secundária dos motivos conservados na região C-terminal das proteínas R2R3-MYB e denominações dos subgrupos das proteínas R2R3-MYB do feijão-caupi (Vu). **(A)** A árvore NJ inclui 64 proteínas R2R3-MYB do feijão-caupi e apresenta 18 subgrupos (C1-C18) marcados com fundos coloridos, para facilitar a identificação da subfamília com alto valor preditivo. Os números ao lado dos ramos representam valores de bootstrap ($\geq 60\%$) com base em 1000 repetições. **(B)** Estrutura secundária dos motivos conservados na região C-terminal das proteínas R2R3-MYB em 18 subgrupos. Os motivos foram detectados utilizando o software MEME e são representados graficamente como caixas desenhadas à escala para uma proteína MYB representante de cada subfamília. Caixas em azul representam o domínio R2R3. *Proteínas VuMYB que não apresentam o padrão de motivo característico do seu subgrupo.

A distribuição dos VuMYBs nos 18 subgrupos não foi uniforme, variando entre 7 (C18) e 1 (C5, C7 e C17) por subgrupo. Em ordem decrescente de tamanho, os subgrupos seguiram a seguinte ordem: C9 e C18 com sete sequências, cada; C12 e C15 com seis, cada; Cinco sequências nos subgrupos C4, C11, C13 e C14; Subgrupos C2 e C10 com quatro sequências; C8 e C16 contendo três MYBs, e; C1, C3, C5, C6, C7 e C17, cada clado com duas sequências MYB. A distribuição dos subgrupos pode ser observada na Figura 3a.

Para os 18 subgrupos formados, 13 continham pelo menos um AtMYB entre os membros, dentre os quais nove já tinham sido previamente classificados nos subgrupos descritos por Strack *et al.* (2001) e Dubos *et al.* (2010). Um dos agrupamentos observados na árvore (VuMYB101, AtMYB16 e AtMYB51) divergiu do reportado anteriormente para *Arabidopsis*, onde tais sequências formavam dois subgrupos distintos (a saber, S9 e S12, respectivamente). Entretanto, tal clado apresentou um baixo suporte de bootstrap, não sendo considerado como subgrupo no presente estudo.

A distribuição das proteínas VuMYB em subgrupos, baseada na similaridade de suas sequências com as sequências de *A. thaliana* utilizadas no presente estudo, permitiu ainda que se fosse atribuída prováveis funções aos subclados agrupados com AtMYBs, uma vez que os estes já possuem funções caracterizadas. De forma geral, observa-se que os genes MYB aqui analisados estão envolvidos no desenvolvimento, metabolismo e defesa das plantas. As possíveis funções designadas de acordo com a presença das sequências AtMYBs, pode ser observada na Tabela 1.

Tabela 1: Funções designadas aos subgrupos correspondentes do agrupamento com as sequências de AtMYB, segundo a classificação designada por Dubos *et al.* (2010).

Subgrupo Correspondente de <i>Arabidopsis</i>	Subgrupo Correspondente do Feijão-caupi	Função	Papel desempenhado
S1	C11	Defesa	Estresse abiótico (seca) / estresse biótico (AS - HR)
S2	C4	Defesa	Estresse abiótico
S4	C2	Metabolismo	Via dos fenilpropanoides
S9	C5	Diferenciação celular	Controle sobre etapas de diferenciação
S13	C9	Metabolismo e desenvolvimento	Deposição e extrusão de mucilagem, Síntese de lignina, fechamento estomático
S18	C13	Desenvolvimento e defesa	Desenvolvimento do estame e antera, resposta a estresse abiótico e hormônio ABA
S21	C14	Metabolismo e desenvolvimento	Espessamento da parede celular e regulação de meristemas axilares
S23	C15	Não determinada	Não determinado
S24	C7	Não determinada	Não determinado

No total, 16 motivos conservados com comprimento variável (18-250 aa) foram detectados na região C-terminal das proteínas MYB do feijão-caupi e *Arabidopsis* (Tabela 2). Notoriamente, a maioria dos membros de um mesmo subgrupo compartilharam um ou mais motivos fora do domínio MYB. À exceção dos motivos 1 e 6, todos os motivos identificados foram únicos para os seus subgrupos. Os motivos 1, 2, 5 e 16 foram encontrados estreitamente adjacentes ao domínio MYB, mas nunca apareceram na mesma sequência (Figura 3b).

Tabela 2: Sequências consensos dos motivos da região C-terminal de proteínas R2R3-MYB do feijão-caupi e Arabidopsis identificados através de análise na ferramenta MEME.

Motivo	Extensão	e-value	Sequência consenso
1	19	2.6e-205	KKKLIQMGIDPVTHRPLTD
2	24	4.9e-074	IRLNSMKNKDRRRPSIHDITTVNNG
3	105	2.2e-072	LVECCKELEEGRHAWAAHKKEAAWRLRRVELQLEAEKACRRREKMEEIEAKIKALRE EQRAALDRIEAEYREQLAGLRDAEAEQKLAEQWTAKHIRLTKFLEQ
4	250	1.2e-063	TRALLANYAQTPGQGMTPMRTPQRTMPAGKGDAMMEAEENLARLRESQTPLLGGENP ELHPSDFTGVTPKKKEIQTNPMLTPSMTPGGAGLTPRIGMTPTRDGFSSMTPKGT PFRDELRIINEDMDMHESAKRERQRREEMRRSLRSLGGLPQPKNEYQIVMQPVPEE AEEPEEKIEEDMSDRIAREKAEQEQALLKKRSKVLQRELPRPPAASLEVIRNSLIR ADGDKSSFVPPPTPIEQADKMIR
5	28	2.7e-034	RIKRRQRAGLPLYPEIYVQALHWKQEY
6	24	2.3e-025	RHMAQWESARLEAEARLSRESYLF
7	99	5.9e-024	RPPPGFYDVAEDRPVEQVKFPTTIELEGKRRVDIEAQLRKQDIKNKIAQRQDAPA AILQANKLNDPEVVRKRSKMLPPPQISDQELEEIAKMGYA
8	18	2.9e-023	EFLAVMQDMIRKEVRSYM
9	80	8.6e-022	PRKLRPGEIDPNPEAKPARPDPVDMDEDEKEMLSEARARLANTKGKKAKRKAREKQ LEEARRLASLQKKRELKAAGIDGR
10	27	5.7e-018	RGQWERRLQTDIHEMAKRALCEALSPDK
11	108	8.0e-017	IEDFEEDEMKEADKMIKEEGKYLCVAMGHENKTLDEFIEAHRTCNDLMYFPTRNAY GLSSVAGNMEKVAAFQEEFENVRKKMEEDKKKMERMKKKYKVYTKGHERRA
12	29	3.7e-016	PPGGHMSAVGTPVMLPPPAHMHYHVHAH
13	21	4.6e-016	YASSAENIARLLKGWMKKSPK
14	79	4.7e-016	MHDSCAMSPHGSSSESVVVENGFGSFLDAYGEVTFENFWMQPYVADMSHVPSELL VPSEVESEYFTPMYDDLWGQSHFY
15	35	2.8e-012	HYMMIPDSHTMTNGNYPTSKPLFGAVKLELPSFY
16	37	3.6e-016	EKQQREKKEINKRIDPIDNSKYDHILESFAEKLKVEH

Perfil de Expressão Transcricional dos VuMYBs

Com o propósito de inferir sobre os possíveis papéis dos candidatos a genes MYB nos mecanismos de defesa do feijão-caupi em resposta a infecção pelo CPSMV, foi analisada a expressão das tags obtidas a partir do alinhamento com os VuMYBs identificados no presente estudo.

Um total de 236 tags SuperSAGE foram retornadas, das quais 93 (39%) mostraram expressão diferencial em pelo menos um tratamento (bulk de 30', 60' e 90' e/ou 16h) com o inóculo viral. As tags superexpressas constituíram um universo de 32 sequências disponíveis para o desenho dos *primers*, das quais dez foram utilizadas para o desenho dos iniciadores (Tabela 3). Em uma triagem inicial para definir a especificidade das sequências VuMYB, dois pares de *primers* (M39756 e M103289) foram eliminados por terem apresentado mais de uma banda na PCR convencional. Por conseguinte, a eficiência de amplificação dos *primers* específicos de VuMYB foi testada em quatro diluições seriadas (concentrado, 1:10, 1:100 e 1:1.000) do bulk de cDNA das amostras tratadas e do controle, obtendo-se ao fim cinco *primers* VuMYB eficientes para a análise de RT-qPCR. Como controle endógeno (*housekeeping*) para o estresse avaliado, foram utilizados no presente estudo a combinação dos genes da Actina, β -tubulina e SKIP (Amorim et al. 'in prep.').

Os resultados dos produtos das RT-qPCRs para cinco genes-alvo e três genes de referência foram registrados pelo programa LineGene 9600. Os dados brutos com os valores de Cq para cada repetição técnica de cada amostra de cDNA foram inseridos no programa REST. Assim, obteve-se uma razão de expressão dos genes alvos para as amostras de cDNAs de cada período de tratamento com o CPSMV (30 min, 60 min e 16 h) em relação ao tratamento controle (30 min e 16 h).

Tabela 3: Tags selecionadas com suas respectivas ESTs utilizadas como *input* para o desenho e amplificação dos *primers* via RT-qPCR.

Tag up-regulada	EST (NordEST)	Proteína VuMYB correspondente	Primers (L / R)
CMV_75892	PO15968.1	VuMYB104	TGGCTCCTCATGGATTCTTC / CTTGGGAATGACTGGGATTG
CMV_39756	PO16969.1	VuMYB11	AGAGTCACAAGCCCAGATGC / CCCACTCCTCCTTAAACCTG
CMV_15835	SI8553.1	VuMYB114	AACCGCTTTAATCTCCAACG / GAAGCCTCTGACACCAAACC
CMV_44427	PO3577.1	VuMYB23	ACACATGGCATAACGGTCTCG / CATCTATGAGCAGATGTGTGTGG
CMV_103289	SI25873.1	VuMYB29	TTGACCCTTCTTCTTCCCTTC / GAACGAACATTGGAGACAGC
CMV_60030	SI16631.2	VuMYB39	CCCACAGATCATCGTACATAGG / GACTTTGGCAGCTTTCTGGA
CMV_38967	PO12503.1	VuMYB78	CCAATGATAACAGCACCACA / TCACTGGAAGCAGTGTGTCAGC
CMV_90151	SI41294.1	Palavra-chave	CTTATGGAGCGGATTCAAGC / CGCATCAGAGAGAGAAAGAGG
CMV_105535	Contig12389	VuMYB101	CCCTTTCTCACCTAGACACACAG / TTCTGGTCTTCTTGAACAGTCC
CMV_70689	SI33856.1	VuMYB18	TGGACTAACAGGGATGATGC / TTTCATGTTGCCACTGTACG

O gene M105525 teve sua expressão regulada negativamente logo no início do estresse, sendo provavelmente alvo do patógeno que conseguiu inibi-lo já no período inicial da infecção. Os genes M75892, M44427 e M38967 tiveram sua expressão induzida com o decorrer do tempo de estresse (16 h) sugerindo que a planta estava iniciando a sua reprogramação transcricional em resposta ao patógeno, sendo o tempo inicial (30 e 60 min) curto, mesmo se tratando de um genótipo tolerante (Figura 4).

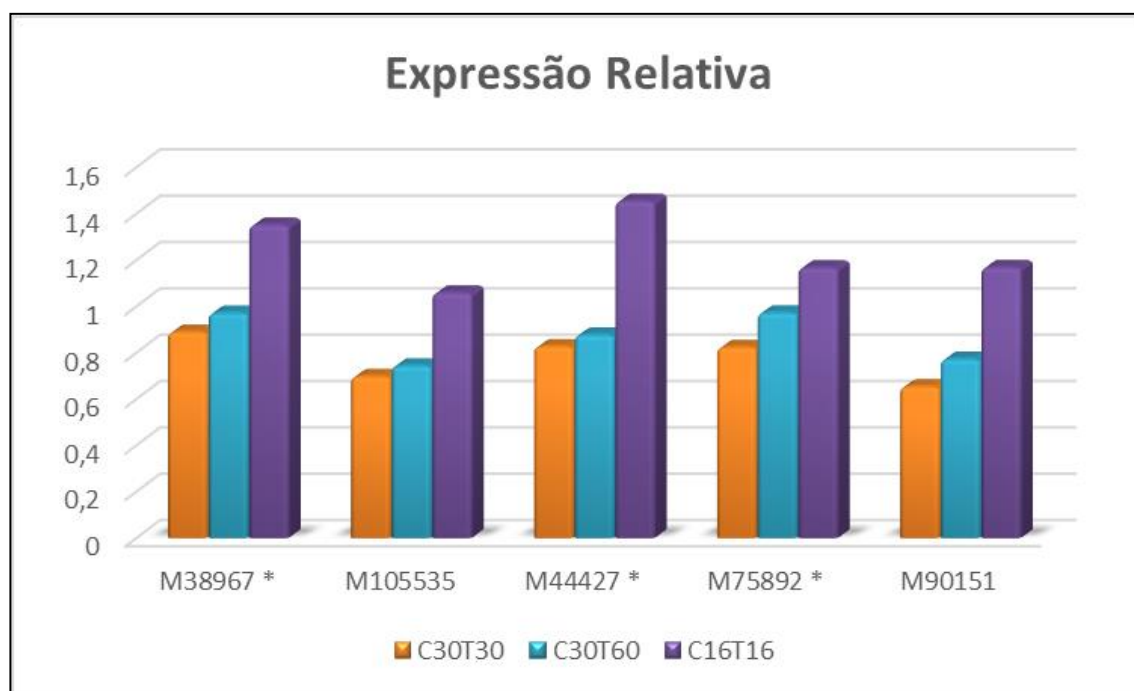


Figura 4: Quantificação relativa dos genes MYB sob diferentes tempos de estresse viral; C30T30 (controle tempo 30 e tratamento com vírus tempo 30); C30T60 (controle tempo 30 e tratamento com vírus tempo 60) e C16T16 (controle tempo 16 horas e tratamento com vírus tempo 16 horas).
* Genes up-regulados em comparação ao controle.

DISCUSSÃO

Sequências transcriptômicas mostram-se como um valioso recurso para espécies que não possuem o genoma completo sequenciado, pois auxiliam na descoberta de novos genes, no desenvolvimento de marcadores moleculares, além de permitirem análises evolutivas e de expressão (Guimarães, *et al.*, 2012). A análise do transcriptoma do feijão-caupi reportou um total de 86 proteínas MYB confirmadas. O grande tamanho desta família gênica também foi confirmado em outras plantas, como *A. thaliana* (Dubos *et al.*, 2010), *G. max* (Du *et al.*, 2012b), *V. vinifera* (Wilkins *et al.*, 2009), *Populus trichocarpa* (Feller *et al.*, 2011) e nas monocotiledôneas *Oryza sativa* (arroz), *Zea mays* (milho) (Yanhui *et al.*, 2006) e

Sorghum bicolor (sorgo) (Jiang *et al.*, 2004), destacando-se como a família mais abundante de TFs em plantas.

A subfamília R2R3-MYB consistiu na classe mais representativa entre as subfamílias MYB identificadas no feijão-caupi, seguido pela subclasse R1-MYB, com 73,25 e 23,25% das proteínas MYB, respectivamente. Proporções similares para a ocorrência destas subfamílias foram reportadas por Cavalcanti-Soares *et al.* (2009), nos transcriptomas do eucalipto, cana-de-açúcar e arroz. Em soja, a subclasse R2R3-MYB representa 33% dos 791 MYBs preditos (Aoyagi *et al.*, 2014). Dada a grande abundância e importância desta subfamília entre as subclasses de MYB, vários estudos têm concentrado suas análises na identificação e caracterização das proteínas R2R3-MYB (Li *et al.*, 2012; Du *et al.*, 2012a; Tombuloglu *et al.*, 2013; Aoyagi *et al.*, 2014). Uma hipótese para a origem desta grande subfamília nos vegetais sugere que tais proteínas tenham se originado pela perda de um domínio R1 em proteínas R3-MYB. Entretanto, este evento evolutivo ainda não é absolutamente claro, havendo outra hipótese de que a subfamília R2R3-MYB tenha se originado por um evento de duplicação do domínio R1 com subsequente ganho do mesmo (Feller *et al.*, 2011). Embora a identificação do conjunto completo de proteínas MYB de feijão-caupi só deva ser possível após a conclusão do sequenciamento completo do genoma, espera-se pouca ou nenhuma redundância nos dados aqui gerados.

As proteínas R2R3-MYB do feijão-caupi mostraram-se bastante conservadas entre as sequências de aa dentro das regiões de domínio MYB. É possível observar no alinhamento dos domínios que na repetição R2 nove resíduos de aa mostraram-se totalmente conservados entre todos os genes MYB analisados, enquanto que apenas quatro foram completamente idênticos no

domínio R3 (Figura 2). Zhang *et al.* (2012) também verificaram níveis similares de conservação de aa nos domínios R2R3 de proteínas MYB do trigo, diferindo das proteínas MYB do feijão-caupi nas posições G-21 e L-52 na repetição R2 e, E-65 na repetição R3. Já Du *et al.* (2012a) identificaram nas proteínas R2R3-MYB de milho, além dos mesmos resíduos conservados que os do feijão-caupi, aa em outras posições com níveis totais de conservação (a saber, R-40, W-49 para repetição R2 e N-94, K-97 e N-98 para a repetição R3 de MYB do milho). Acredita-se que tais variações entre esses dois últimos organismos tenham surgido depois da divergência entre monocotiledôneas e dicotiledôneas, visto que a amplificação das proteínas MYB ocorreu antes desse evento (Jiang *et al.*, 2004).

Uma proteína MYB é caracterizada por conter três resíduos de W espaçados regularmente entre si em cada repetição R, os quais possivelmente desempenham papéis significativos na conformação tridimensional da proteína e na interação com sequências específicas do DNA (Dubos *et al.*, 2010). Tal característica pode ser observada nas proteínas VuMYB para a repetição R2, porém, o primeiro resíduo de W na repetição R3 foi substituído em sua totalidade por uma Fenilalanina (F) e em menor frequência por uma Leucina (L) ou Isoleucina (I). Esta substituição é comumente observada nas proteínas MYB de plantas (Stracke *et al.*, 2001; Jia *et al.*, 2004; Zhang *et al.*, 2012; Du *et al.*, 2012a, 2013) e não afeta a conformação das proteínas devido aos aa substituídos serem predominantemente hidrofóbicos (com grupo hidrocarboneto aromático).

O terceiro resíduo de W das repetições R2 e R3 apresentou-se conservado em todos os membros, exceto para a proteína VuMYB1, no qual foi substituído por uma Fenilalanina (F), sendo que as proteínas VuMYB61 e VuMYB_SI69686

apresentaram uma Tirosina (Y) no lugar do terceiro resíduo de W da repetição R3. É notório entre as repetições do domínio MYB que a terceira hélice apresenta-se como a mais conservada entre as α -hélices. Os principais resíduos conservados neste domínio foram distribuídos principalmente entre o segundo e terceiro Trp conservados em cada repetição. Assim, a primeira parte de cada repetição do domínio MYB foi aparentemente menos conservada. Consistente com isso, já foi demonstrada sua importância no reconhecimento dos elementos cis de genes-alvo para desempenhar atividade de ligação ao DNA (Ogata *et al.*, 1994; Du *et al.*, 2013). Dessa forma, alterações de resíduos nesta região específica podem tanto afetar na interação da proteína com o DNA como resultar no reconhecimento de novos genes (Jiang *et al.*, 2003; Du *et al.*, 2012b).

De maneira geral, inserções ou deleções não são comuns nos domínios MYB. Du *et al.* (2012b) também identificaram em aproximadamente 6% das 244 R2R3-MYB de soja, e associaram tal resultado ao *splicing* alternativo das proteínas. Por outro lado, Jiang *et al.* (2003) reportaram uma inserção no gene MYB Zm19072746 do milho na repetição R3, associando-o a um alelo não funcional. As proteínas MYB do feijão-caupi apresentaram uma elevada taxa de *indels* na segunda α -hélice da repetição R2. O efeito de tais mutações nessas proteínas ainda é desconhecido. Entretanto sabe-se que tal hélice, juntamente com a primeira, está mais sujeita a acomodar alterações de aa sem efeitos deletérios (Jia *et al.*, 2004).

A ancoragem das sequências VuMYBs nos cromossomos virtuais do feijão-comum, revelou que tais proteínas se distribuem de modo desigual dentro do genoma. Esse padrão de distribuição irregular dos genes MYB é comum entre os vegetais e já tem sido observado em vários outros organismos, como *Arabidopsis*,

O. sativa (Katiyar *et al.*, 2012), *Cucumis sativus* (Li *et al.*, 2012), *Hordeum vulgare* (Tombuloglu *et al.*, 2013) e *Setaria itálica* (Muthamilarasam *et al.*, 2014).

Eventos evolutivos, como duplicação gênica e poliploidização têm sido reconhecidos como parte do processo evolutivo das plantas, contribuindo desta forma para origem de novas funções gênicas, subjacente às origens da novidade evolutiva (Kent *et al.*, 2003; Cannon *et al.*, 2004). Feller *et al.* (2011) estudaram a evolução dos genes MYB em plantas e verificaram que a grande diversidade dessa família é decorrente de duplicações em tandem de grandes segmentos cromossômicos e dos processos de poliploidização dos genomas, como é o caso de *Arabidopsis* (Cannon *et al.*, 2004), *Populus trichocarpa* (Wilkins *et al.*, 2009) e soja (Du *et al.*, 2012b). No presente estudo, foi observada a ocorrência de duplicações em tandem seguidas ou não de translocação intercromossômica das proteínas VuMYB e acredita-se que tais eventos evolutivos tenham contribuído para o aumento do número de genes desta família.

É provável que diferentes proteínas MYBs tenham evoluído pela duplicação do seu domínio de ligação ao DNA seguido por uma série de mutações sinônimas e/ou não-sinônimas (especialmente na região C-terminal) gerando novas funções durante o processo evolutivo (Du *et al.*, 2012b). De maneira geral, os resultados aqui obtidos indicam que a duplicação em tandem e a translocação intercromossômica contribuíram para a expansão da família MYB no feijão-caupi. O número de duplicações em tandem foi menor que o de duplicações de segmentos seguidas de translocação. Portanto, este último processo pode ter sido o principal contribuinte para a expansão de sequências MYB no feijão-caupi.

As relações estruturais e funcionais entre as proteínas R2R3-MYB do feijão-caupi e *Arabidopsis* foram determinadas pela análise fenética utilizando o

método de NJ. Os baixos valores de suporte obtido nos ramos internos da árvore estão em consonância com o observado em outros organismos, provavelmente devido ao tamanho relativamente curto do domínio MYB, com poucas posições de carácter informativo. Em contrapartida, os valores de bootstrap dos ramos mais externos que definem os grupos foram mais altos.

De modo geral, as estruturas e funções de genes agrupados em um mesmo subclado aparecem conservadas entre as espécies de plantas. Logo, o conhecimento das funções de alguns dos membros de um subgrupo deve facilitar a confirmação das relações funcionais dos seus ortólogos (Du *et al.*, 2012a). Consistente com isso, a análise dos motivos conservados na região C-terminal, responsáveis pela atividade da proteína na regulação transcricional (Feller *et al.*, 2011) confirma os grupos formados pela análise fenética. Assim, embora as informações que ligam cada um destes motivos a funções exatas sejam escassas, a partilha destas estruturas por integrantes de um mesmo cluster fornece um meio para inferir funções putativas com base na função da seed sequence de *A. thaliana* (Aoyagi *et al.*, 2014).

A distribuição dos 16 motivos identificados no presente estudo entre os subclados formados nos permitiu verificar que apenas os motivos 1 e 6 não foram exclusivos de apenas um subgrupo. Curiosamente, os motivos 1, 2, 5 e 16 foram encontrados em posições estreitamente adjacentes ao domínio MYB, mas nunca apareceram numa mesma sequência. De acordo com a análise da estrutura de proteínas MYB, esta localização específica deve desempenhar um papel na ativação (Cai *et al.*, 2012).

Foi comum encontrar que duas ou mais sequências VuMYBs formavam um subgrupo com uma única proteína de *Arabidopsis*. Por exemplo, o subgrupo C4 e

C11 incluem representantes AtMYBs envolvidos na defesa contra estresses ambientais (Cominelli *et al.*, 2005; Reyes e Chua, 2007), ao passo que os AtMYBs dos subgrupos C9 e C14 atuam no metabolismo e desenvolvimento das plantas (Newman *et al.*, 2004; Stracke *et al.*, 2007). Notoriamente, as proteínas VuMYBs dos subgrupos C9 e C11 compartilham os mesmos motivos funcionais que seus ortólogos em *Arabidopsis*, ao passo que no subgrupo C14 nenhum dos representantes apresentam motivos funcionais. Assim, os TFs do feijão-caupi agrupados juntamente com os AtMYBs, apesar de não possuírem funções conhecidas compartilham a estrutura semelhante do domínio de ligação, além de motivos encontrados caracteristicamente em tais grupos e, possivelmente, podem desempenhar o mesmo controle sobre as respostas de defesa, desenvolvimento e metabolismo.

Os subclados formados apenas por proteínas VuMYB sugerem que tais sequências sejam específicas da espécie, ou que foram perdidas durante a evolução de *Arabidopsis*, ou ainda adquiridos nas linhagens do feijão-caupi após a divergência a partir do ancestral comum mais recente. Interessantemente, o subclado C10 agrupou metade das proteínas que apresentaram *indels* no seu domínio de ligação ao DNA, assim como também não exibiram a presença de nenhum motivo conservado. Desta forma, não é conclusivo se tais proteínas são exclusivas do feijão-caupi ou se foi apenas um erro de sequenciamento.

Embora as funções da maior parte dos VuMYBs permanecem por ser elucidados, é provável que os membros de um dado subgrupo possam ter origem evolutiva comum e recente, como também uma função conservada.

TFs MYB já foram descritos por atuarem regulando a expressão gênica em resposta a diversos sinais e no controle de processos específicos da planta, tais

como o metabolismo, desenvolvimento e respostas aos estresses bióticos e abióticos (Du *et al.*, 2009; Dubos *et al.*, 2010). Com base nos dados de SuperSAGE, foram selecionados tags de MYBs diferencialmente expressos para serem validados via RT-qPCR frente à infecção pelo CPSMV.

Reforçando a potencial contribuição destes FTs para a defesa no feijão-caupi durante estresse biótico, foi observada a expressão diferencial dos VuMYBs em bioensaio de SuperSAGE com outra virose que ataca a cultura, o CABMV (*Cowpea Aphid-Borne Mosaic Virus*). Dos 91 VuMYBs diferencialmente expressos no feijão-caupi em resposta ao CPSMV, 62 (26%) foram encontradas na biblioteca de Tags do CABMV, sendo que destes apenas um não apresentou expressão diferencial em nenhuma comparação (dados não mostrados).

A proteína VuMYB_78, representativa para tag M38967, a qual teve sua expressão induzida 16 horas após a inoculação do vírus, foi agrupada no subclado C11 juntamente com AtMYB60. Este regulador foi descrito por Cominelli *et al.* (2005) controlando a cascata de sinalização mediada por ABA (ácido abscísico) para regular a abertura e fechamento dos estômatos como resposta à restrição hídrica em *Arabidopsis*. AtMYB60 foi classificado por Dubos *et al.* (2010) no subgrupo S1, juntamente com o AtMYB96 e outras proteínas. Em *Arabidopsis* o gene AtMYB96 atua através da sinalização que integra ABA e sinais de AS (ácido salicílico) para respostas ao estresse hídrico e na resistência a doenças (Seo *et al.*, 2009; Seo e Park, 2010). O hormônio ABA atua sob condições de estresse biótico e abiótico tanto de forma isolada quanto em sinalização cruzada (*crosstalk*), refletindo a importância do equilíbrio hormonal na sobrevivência da planta sob condições de estresse (Seo e Park, 2010). Estes dados levam-nos a inferir que os membros pertencentes a este subgrupo estão propensos a atuarem

em respostas cruzadas (estresse biótico e abiótico), salientando-se que tais proteínas multifuncionais são significativamente importantes para a sobrevivência das plantas sob condições de estresse extremo, uma vez que tais organismos estão sujeitos a estresses múltiplos simultaneamente (Xie *et al.*, 2014).

O subgrupo C18, formado exclusivamente por sequências MYB do feijão-caupi, contém a proteína VuMYB_23, a qual alinhou com a tag M44427 up-regulada em 16 horas após a infecção. Baseado na similaridade dos resíduos no domínio MYB, bem como dos motivos conservados na região C-terminal, a anotação funcional deste gene nos permite inferir que as proteínas agrupadas no subclado C18, juntamente com VuMYB_23, possivelmente estão envolvidas na resposta a infecção por patógenos no feijão-caupi, tornando-os potenciais alvos para estudos funcionais futuros.

A tag M75892, representativa da sequência VuMYB_104, uma proteína da subfamília R1-MYB, foi regulada positivamente 16 horas após a inoculação do vírus. Comparado com TFs MYB da subfamília R2R3, estudos sobre as relações funcionais das proteínas R1-MYB em plantas são ainda limitados (Li *et al.*, 2014). Evidências crescentes sugerem que esta subclasse proteica atua na manutenção da estrutura cromossômica e na resposta ao ciclo de luz nas plantas (Xiong *et al.*, 2014). Em resposta ao estresse abiótico, já foram mostrados que proteínas R1-MYB podem estar envolvidas na tolerância à seca (Shin *et al.*, 2011), salinidade (Ganesan *et al.*, 2012) e frio (Su *et al.*, 2010) em batata, tabaco e arroz, respectivamente. Na soja e no milho, genes R1-MYB foram identificados sendo diferencialmente expressos em resposta ao ataque de agentes patogênicos (Du *et al.*, 2013).

Em síntese, nossos dados mostram que transcritos das duas maiores subclasses de TF MYB estão envolvidas na resposta à infecção pelo CPSMV no feijão-caupi.

AGRADECIMENTOS

Os autores agradecem ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco) e CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte financeiro, e a EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) e IPA (Instituto Agrônômico de Pernambuco) pela concessão do material genético e parceria.

REFERÊNCIAS

- Altschul SF, Thomas LM, Alejandro AS, Jinghui Z, Zheng Z, Webb M and David JL (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25:3389-3402.
- Ambawat S, Sharma P, Yadav NR and Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants* 19(3): 307–321.
- Aoyagi LN, Caitar VSL, Carvalho MCCG, Darben LM, Podanosqui AP, Kuwahara MK, Nepomuceno AL, Abdelnoor RV and Guimarães FCM (2014) Genomic and transcriptomic characterization of the transcription factor family R2R3-MYB in soybean and its involvement in the resistance responses to *Phakopsora pachyrhizi*. *Plant Science* 229: 32–42.
- Bailey TL, Williams N, Misleh C and Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34:W369–W373.
- Booker HM, Umaharan P and McDavid CR (2005) Effect of Cowpea severe mosaic virus on crop growth characteristics and yield of cowpea. *Plant Dis* 89: 515-520.
- Cai H, Tian S and Dong H (2012) Large Scale In Silico Identification of MYB Family Genes from Wheat Expressed Sequence Tags. *Mol Biotechnol* 52:184–192.
- Cannon SB, Mitra A, Baumgarten A, Young ND and May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol* 4:10.
- Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL and Tonelli (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *C Curr Biol* 15(13):1196-200.
- Daniel-Silva M, Oliveira-Silva RL, Ferreira-Neto JRC, Guimarães ACR, Veiga DT, Chabregas SM, Burnquist WL, Kahl G, Benko-Iseppon AM and Kido EA (2013) Expression Analysis of Sugarcane Aquaporin Genes under Water Deficit. *J Nucl Acids* 14.
- Du H, Feng B, Yang S, Huang Y and Tang Y (2012a) The R2R3-MYB Transcription Factor Gene Family in Maize. *BMC Plant Biol* 7: 1-12.
- Du H, Wang YB, Xie Y, Liang Z, Jiang SL, Zhang SS, Huang YB and Tang YX (2013) Genome-Wide Identification and Evolutionary and Expression Analyses of MYB-Related Genes in Land Plants. *DNA Res* 1-12.
- Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB and Tang YX (2012b) Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol* 12:106.
- Du H, Zhang L, Liu L, Tang X, Yang W, Wu Y, Huang Y, and Tang Y (2009) Biochemical and Molecular Characterization of Plant MYB Transcription Factor Family. *Biochemistry* 74(1): 1-11.
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C and Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 15(10): 573-581.
- Feller A, Machemer K, Braun EL and Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* 66: 94-116.

- Ganesan G, Sankararamasubramanian HM, Harikrishnan M, Ashwin G and Parida A (2012) A MYB transcription factor from the grey mangrove is induced by stress and confers NaCl tolerance in tobacco. *J Exp Bot* 63: 4549–4561.
- Guimarães PM, Brasileiro AC, Morgante CV, Martins AC, Pappas G, Silva Jr OB, Togawa R, Soraya CM Leal-Bertioli SC, Ana CG Araujo, Moretzsohn MC, Bertioli DJ, *et al.* (2012) Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. *BMC Genomics* 13(387): 1-15.
- Hou XJ, Li SB, Liu SR, Hu CG and Zhang JZ (2014) Genome-Wide Classification and Evolutionary and Expression Analyses of Citrus MYB Transcription Factor Families in Sweet Orange. *PLoS ONE* 9(11): e112375.
- Jia L, Clegg MT and Jiang T (2004) Evolutionary Dynamics of the DNA-Binding Domains in Putative R2R3-MYB Genes Identified from Rice Subspecies *indica* and *japonica* Genomes. *Plant Physiol* 134: 575-585.
- Jiang C, Gu J, Chopra S, Gu X and Peterson T (2003) Ordered origin of the typical two- and three-repeat MYB genes. *Gene* 326: 13–22.
- Jiang C, Gu X and Peterson T (2004) Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. *Genome Biol* 5: R46.
- Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V and Bansal KC (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genomics* 13(544): 1-19.
- Kent WJ, Baertsch R, Hinrichs A, Miller W and Haussler (2003) D: Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci USA* 100:11484–11489.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ and Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639-1645.
- Li Q, Zhang C, Li J, Wang L and Ren Z (2012) Genome-Wide Identification and Characterization of R2R3MYB Family in *Cucumis sativus*. *PLoS ONE* 7(10): e47576.
- Li HL, Guo D and Peng SQ (2014) Molecular characterization of the *Jatropha curcas* JcR1MYB1 gene encoding a putative R1-MYB transcription factor. *Gen Mol Biol* 37(3): 549-555.
- Libault M, Wan J, Czechowski T, Udvardi M and Stacey G (2007) Identification of 118 *Arabidopsis* transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant –defense elicitor. *Mol Plant Microbe In* 20: 900-91.
- Molina C, Rotter B, Horres R, Udupa SM, Besser B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G and Winter P (2008) SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. *BMC Genomics* 9:553.
- Muthamilarasan M, Khandelwal R, Yadav CB, Bonthala VS, Khan Y and Prasad M (2014) Identification and Molecular Characterization of MYB Transcription Factor Superfamily in C4 Model Plant Foxtail Millet (*Setaria italica* L.). *PLoS ONE* 9(10): e109920.
- Newman LJ, Perazza DE, Juda L and Campbell MM (2004) Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the *det3* mutant phenotype. *Plant J* 37: 239–250.
- Ogata K, Morikawa S, Nakamura H, Sekikawa, A, Inoue T, Kanai H, Sarai A, Ishii S and Nishimura Y (1994) Solution structure of a specific DNA complex of

- the Myb DNA-binding domain with cooperative recognition helices. *Cell* 79: 639–648.
- Pfaffl MW, Horgan GW and Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucl Acids Res* 30: 1-10.
- Reyes JL and Chua NH (2007) ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J* 49: 592–606.
- Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, Varhol RJ, Fjell C, Marra M, Jones S and Siddiqui A (2007) Discovery-Space: an interactive data analysis application. *Genome Biol* 8: R6.
- Seo PJ and Park CM (2010) MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in *Arabidopsis*. *New Phytol* 186: 471–483.
- Seo PJ, Xiang F, Qiao M, Park JY, Lee YN, Kim SG, Lee YH, Park WJ and Park CM (2009) The MYB96 Transcription Factor Mediates Absciscic Acid Signaling during Drought Stress Response in *Arabidopsis*. *Plant Physiol* 151: 275–289.
- Soares-Cavalcanti NM, Wanderley-Nogueira AC, Belarmino LC, Barros PS and Benko-Iseppon AM (2009) Comparative in silico evaluation of MYB transcription factors in eucalyptus, sugarcane and rice transcriptomes. *Lect Notes Comput Sc* 5488: 44-55.
- Stracke R, Werber M and Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* 4: 447-456.
- Stracke R, Ishihara H, Huep G, Barsch A, Mehrstens F, Niehaus K and Weisshaar B (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J* 50: 660–677
- Su CF, Wang YC, Hsieh TH, Lu CA, Tseng TH and Yu SM (2010) A novel MYBS3-dependent pathway confers cold tolerance in rice. *Plant Physiol* 153: 145–158.
- Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30: 2725-2729.
- Tombuloglu H, Kekec G, Sakcali MS and Turgay Unver (2013) Transcriptome-wide identification of R2R3-MYB transcription factors in barley with their boron responsive expression analysis. *Mol Genet Genomics* 288(3-4): 141-155.
- Wilkins O, Nahal H, Foong J, Provart NJ and Campbell MM (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Phy* 149: 981–993.
- Xie RJ, Zheng L, Deng L, He SL, Yi SL, Lv Q and Zheng YQ (2014) The Role Of R2R3MYB Transcription Factors In Plant Stress Tolerance. *J Anim Plant Sci* 24(6): 1821-1833.
- Xiong H, Li J, Liu P, Duan J, Zhao Y, Xiao Guo X, Li Y, Zhang H, Ali J and Li Z (2014) Overexpression of OsMYB48-1, a Novel MYB-Related Transcription Factor, Enhances Drought and Salinity Tolerance in Rice. *PLoS ONE* 9(3): e92913.
- Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, et al. (2006) The MYB transcription factor

- superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol* 60: 107–124.
- Zhang H, Jin JP, Tang L, Zhao Y, Gu XC, Gao G and Luo JC (2011) PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res* 39: D1114-D1117.
- Zhang L, Zhao G, Jia J, Liu X and Kong X (2012) Molecular characterization of 60 isolated wheat MYB genes and analysis of their expression during abiotic stress. *J Exp Bot* 63(1): 203–214.
- Zhou B, Jia Z, Tian S, Wang X, Gou Z, Lü B and Dong H (2013) AtMYB44 positively modulates disease resistance to *Pseudomonas syringae* through the salicylic acid signaling pathway in *Arabidopsis*. *Funct Plant Biol* 40:304-313.

Int. J. Mol. Sci. **2015**, *16*, 1-x manuscripts

OPEN ACCESS

International Journal of
Molecular Sciences

ISSN 1422-0067

www.mdpi.com/journal/ijms

Type of the Paper: Article.

Selection of reliable reference genes for expression evaluation during abiotic and biotic stresses in cowpea (*Vigna unguiculata*)

Lidiane L. B. Amorim^{1,2}, Mitalle K. S. Matos^{1,†}, Flávia Tadeu Araújo^{1,†}, José R. C. Ferreira-Neto^{1,†}, João P. Bezerra-Neto^{1,†}, Valesca Pandolfi^{1,†}, Ederson Akio-Kido^{1,†}, Ana M. Benko-Iseppon^{1,*†} and Mauro Guida-Santos^{3,†}

¹Federal University of Pernambuco, Center of Biological Sciences, Av. Prof. Moraes Rego 1235, CEP 50.670-420, Recife, PE, Brazil; E-Mails: mitally_karen@hotmail.com; flaviaaraujo_8@hotmail.com; netocostaferreira@gmail.com; pacifico.joao@gmail.com; valescapandolfi@gmail.com; ederson.kido@gmail.com;

²Federal Institute of Education, Science and Technology of Piauí, Campus Oeiras, CEP 64.500-000, Oeiras, PI, Brazil; E-Mail: lidiane.amorim@gmail.com;

³Federal University of Pernambuco, Botany Department, Center of Biological Sciences, Av. Prof. Moraes Rego s/n, CEP 50.670-901, Recife, PE, Brazil; E-Mail: mauroguida@yahoo.com.br;

† *These authors contributed equally to this work.*

* Author to whom correspondence should be addressed; E-Mail: ana.benko.iseppon@pq.cnpq;
Tel.: +55-81-2126-7816; Fax: +55-82-2126-8522.

Received: / Accepted: / Published:

Abstract: Cowpea is an agricultural crop of significant economic importance. However, environmental degradation impairs cowpeas growth and production.

Gene-associated expression by 2nd generation sequencing, as HT-SuperSAGE (High Throughput Super Serial Analysis of Gene Expression) can provide insights into mechanisms of tolerance to abiotic stress. Quantitative real-time PCR (RT-qPCR), a sensitive and accurate method to estimate transcript abundance expression, depends on the reliable use of reference genes (RGs) for accurate data normalization. Until now, no work was performed to evaluate expression changes of a significant set of commonly used RGs in cowpea under abiotic and biotic stress. Using two distinct algorithms implemented by geNorm and NormFinder we evaluated a total of seven RGs (*Ubq*, *Ubq10*, *EF1- α* , *Act2/7*, *β -Tub*, *Ukn1*, *Skip16*) in three different experimental sets (control versus drought-stressed roots, control versus salt-stressed roots, and control versus virus-stressed leaves). NormFinder and geNorm revealed that *β -tubulin* was the most stable gene under drought stress, while *Skip16* was most stable under salinity stress. For the Cowpea severe mosaic virus treatment, *Act2/7* and *Skip16* were the best scoring genes. Our results globally indicated *EF1- α* as the gene whose expression is less stable among all RG-candidates limiting its use as an internal control in cowpea. Moreover, differences were evident in the optimal gene combination of RGs obtained by both geNorm and NormFinder. The relative expression level of three cowpea genes (*LTP*, *CHI* and *CHS*) was analyzed in order to validate the best combination of RGs selected for this study. Our findings can be used as a guideline for future works on gene expression in cowpea under different conditions.

Keywords: RT-qPCR; normalization; expression assay; root tissue; salinity; drought.

1. Introduction

Gene expression level in plants may be affected by environmental factors. Environmental stresses, such as cold, drought, and contamination of soils by high salt concentration, are major factors that limit plant growth and development [1]. As a consequence of a combination of different abiotic stresses, there is often a reduction in agricultural yield resulting in considerable economic impact for producers and consumers [2,3].

Plant responses to abiotic stresses are very complex and involve changes in morphological, physiological, biochemical and transcriptome levels through activation or deactivation of a large number of genes [4]. Understanding the functions of such stress-induced genes with high-throughput technologies and the analysis of new genomic data may provide insights into complex mechanisms of stress tolerance and could lead to novel and superior breeding strategies [5]. Taking advantage of the knowledge gained from the study of three model legumes genomes (*Medicago truncatula* L., *Lotus japonicas* L. and *Glycine max* L.) efforts are currently being focused on cowpea (*Vigna unguiculata* (L.) Walp.) with the framework of new Brazilian and international initiatives.

Cowpea is a multipurpose legume cultivated in semi-arid regions of Africa, Brazil and United States by subsistence farmers for either human and livestock consumption [6,7]. In Brazil, cowpea is very important in the North and Northeastern regions, which have a tradition in its cultivation, trade and consumption [8]. Cowpea is an excellent system for investigating the genetic basis of drought tolerance because of its ability to fix nitrogen in poor soil and greater drought tolerance when compared to other legumes as common bean and soybean [9, 10].

Improving drought and salinity tolerance is a priority for Brazilian cowpea breeding programs. To this end, we have developed the Brazilian cowpea transcriptomic database (NordEST) integrating the results achieved by different methodologies such as expressed sequence tag library (EST), High Throughput HT-SuperSAGE and RNA-seq. NordEST provides a workbench for measuring gene expression level, gene interactions and their localization and physical interaction network, allowing the identification of target genes and metabolic pathways, vital for the development and study of plants tolerant to challenging

situation. Estimation of the transcript abundance and validation of the genes found to be involved in drought and salinity response via real-time quantitative polymerase chain reaction (RT-qPCR) have been crucial to on-going studies.

RT-qPCR is the foremost method choice for detection of low abundant mRNAs and slight variation in gene expression due to its high sensitivity, reliable specificity, simplicity and cost-effectiveness [11]. However, several variables must be considered when using RT-qPCR to analyze gene expression, including the amount of starting material, RNA recovery and integrity, amplification efficiencies, and the identification of stable reference genes or endogenous control. Of these, the use of suitable reference genes (RGs) for the accurate normalization of the transcript expression level of a target gene is an elementary prerequisite for reliable results in any RT-qPCR analysis [12]. Reference genes should ideally be constitutively expressed in a wide variety of tissues or cell types and should not be affected by the experimental procedure and treatments introduced. Since reference genes are required for cellular survival, it is assumed that they are stably expressed and are often used without validating their suitability [13,14]. Nevertheless, increasing evidences show that the transcription level of commonly used reference genes, such as *18s rRNA*, *25s rRNA*, *GADPH*, *elongation factor 1-alpha*, *actin* and *polyubiquitin* can vary considerably depending on the state of development and under variable physiological conditions, resulting in a misinterpretation of the performance of the target gene, and thus, their choice is essential before initiating a gene expression analysis. To date, a universal reference gene has not been identified [15,16].

Fortunately, various statistical methods have been developed to validate the expression stability of potential reference genes, and to allow the selection of the most suitable gene under particular experimental condition [17]. Normfind and geNorm are the most widely used free software programs [18,19].

For cowpea, only a small number of validated reference genes are available to date. Many of these studies have used 18S ribosomal RNA (18S rRNA), 26S rRNA or Actin as a reference genes [20,21,22], so a comparison of additional potential reference genes in cowpea has not yet been reported. In this context, the present study was undertaken to select and validate additional suitable reference genes for effective normalization of the RT-qPCR data in cowpea. Additionally, this is the first systematic study to be conducted on the expression stability of

reference genes in cowpea, under two different abiotic stress conditions, drought and salinity.

2. Results and Discussion

2.1. Efficiency of the primers for each reference gene

PCR efficiency is calculated by quantifying the increase of amplified product after each thermocycle in samples with a range of transcript abundance, using different cDNA concentrations [23]. Factors known to affect PCR efficiency include template quality (cDNA or DNA), low expression of the gene, primer design (unspecific PCR products), and cycle conditions. The rate of amplification efficiency at 100 % or close to it indicates that the reaction conditions are optimal and the results obtained should be highly reproducible [24].

All PCR assays produced a single amplicon, as shown by the absence of nonspecific bands in electrophoresis analysis and the presence of a single sharp peak in melting curve analysis. The PCR efficiency of each reference gene, which ranged from 91% for *EF1- α* to 109% for *Act2/7* and *Skip16* is shown in Table 1, and these efficiencies were used in the calculations for the subsequent expression data. All PCR primer pairs showed correlation coefficients of $R^2 > 0.98$ (the nearer to 1.0 the better).

Table 1. Gene name, slopes, amplification efficiency (E) and correlation efficiency (R^2) of each reference gene analyzed in this study.

Name	Gene description	Literature	Slope	Efficiency	Y	R^2
<i>Act2/7</i>	Actin 2/7	-	-3.1	109%	43.28	-0.995
<i>β-Tub</i>	β -tubulin	Eticha <i>et al.</i> [25]	-3.3	99%	35.34	-0.994
<i>EF1-α</i>	Elongation factor 1- α	Guerrero-González <i>et al.</i> [26]	-3.5	91%	33.12	-0.999
<i>Skip16</i>	SKP1/ASK-interacting protein 16	Borges <i>et al.</i> [27]	-3.1	109%	35.77	-0.996
<i>Ubq</i>	Ubiquitin-conjugating	Borges <i>et al.</i> [27]	-3.3	99%	34.63	-0.998
<i>Ubq10</i>	Polyubiquitin 10	-	-3.4	96%	40.49	-0.989
<i>Ukn1</i>	Unknown	Borges <i>et al.</i> [27]	-3.3	99%	35.17	-0.989

2.2. Cowpea candidate reference genes stability rankings during different treatments according to GeNorm and NormFinder

The gene expression stabilities of seven reference genes were analyzed in cowpea under drought and salinity using geNorm and NormFinder software. For inoculated with cowpea severe mosaic, six reference genes were available. The geNorm application for Microsoft Excel provides a measure of gene expression stability of a list of selected reference genes in all samples, and ranking them according to gene stability measurements (M) [28]. The default limit value suggested by GeNorm is $M = 1.5$, the reference gene, of which the M value is less than 1.5, is considered to be used as reference gene [29].

In this study, the seven reference genes showed great expression stabilities under drought and salinity abiotic stress conditions because all the M values were below the threshold of 1.5. Under drought treatment, the most stably expressed gene was *β -tubulin* ($M = 0.460$), the most unstable was *EF1- α* ($M = 0.781$), whereas *Act2/7 + Ubq* was the best gene combination (Figure 1a). Under salt treatment, geNorm identified *Skip16* ($M = 0.494$) as the best candidate reference gene, while *EF1- α* was found to be the least stable one ($M = 0.976$) and *Ukn1 + Ubq* was the best gene combination (Figure 1b). For late cowpea severe mosaic

virus exposure, the six reference genes showed great expression stabilities and GeNorm ranked *Act2/7* and *β -tubulin* simultaneously as the most stable genes, and the *Skip16* as the third best gene (Figure 1c).

NormFinder ranked, for the drought stress treatment, *β -tubulin* as the best gene with stability (stability value = 0.104) while the *EF1- α* resulted the least stable (stability value = 0.227), in accordance with the result from geNorm. Furthermore, the *Ukn1* and *Ubq10* was the best gene combination (stability value = 0.080) in disagreement with the geNorm outputs (Figure 1d). For the salinity stress, *Skip16* was ranked as the most stable gene (stability value = 0.099) whereas the *EF1- α* was the least stable (stability values = 0.324), in agreement with the result from geNorm. Additionally, *Ubq* and *Skip16* were recommended as the best combination with a stability value of 0.072, in divergence with the geNorm outputs (Figure 1e). For cowpea virus assay, NormFinder ranked *Ubq10* in the first position (stability value = 0.279) and *Act2/7* and *Skip16* were selected as the second and the third best genes, respectively (stability value = 0.313 and 0.324, respectively).

These discrepancies between the results of the best combination of two genes in the same treatment should not be surprising since the two softwares are based on distinct calculation methods, and similar situations have been reported in many previous studies [30,31]. For example, in pepper, the *Actin* and *GADPH* (Glyceraldehyde-3-phosphate dehydrogenase) exhibited the most stable expression in leaves under salt and drought treatment by geNorm, but the most stable reference gene according to NormFinder was *PPR1* protein [32]. Another interesting result that deserves attention is the poor stability of *EF1- α* gene in the two different treatments (drought and salinity) currently applied. *EF1- α* is an important component for protein biosynthesis [33] and has been reported to be stable in various crops species, such as *Vigna mungo*, coffee, *potato* and *banana* during salinity stress [34-37]. On the other hand, Guerrero-González *et al.* [26] normalized biotic-related RT-qPCR data in common bean with a single *EF1- α* . Also in the legume crop *Caragana intermedia* under various abiotic stresses, *EF1- α* showed good stability and in combination with *SAND family protein* used to normalize the expression of *DREB1* transcription factor [38].

Usually, reference genes include structural genes of the cytoskeleton such as actin and alpha/beta-tubulin [39]. Paralogous of actin genes, such as *Act1*, *Act11*,

Act2 have been ranked at top as reference genes in upland cotton, rice and peanut under drought and salinity [40-42]. In this study, geNorm indicated that *Act2/7* was stable in all experimental conditions. In contrast, *Act2/7* genes have been excluded as good candidate reference genes in soybean tissues at various development stages [43]. β -tubulin is also reported as a significant reference gene for gene expression analysis. Our results would also suggest that these two genes are among the most stable for expression profiling under drought in cowpea.

When comparing the different optimal lists obtained for each treatment, we can also observe that, for all the three stresses involving cowpea, despite the scoring differences, *Act2/7* and *Skip16* are consistently ranked within the most stable genes.

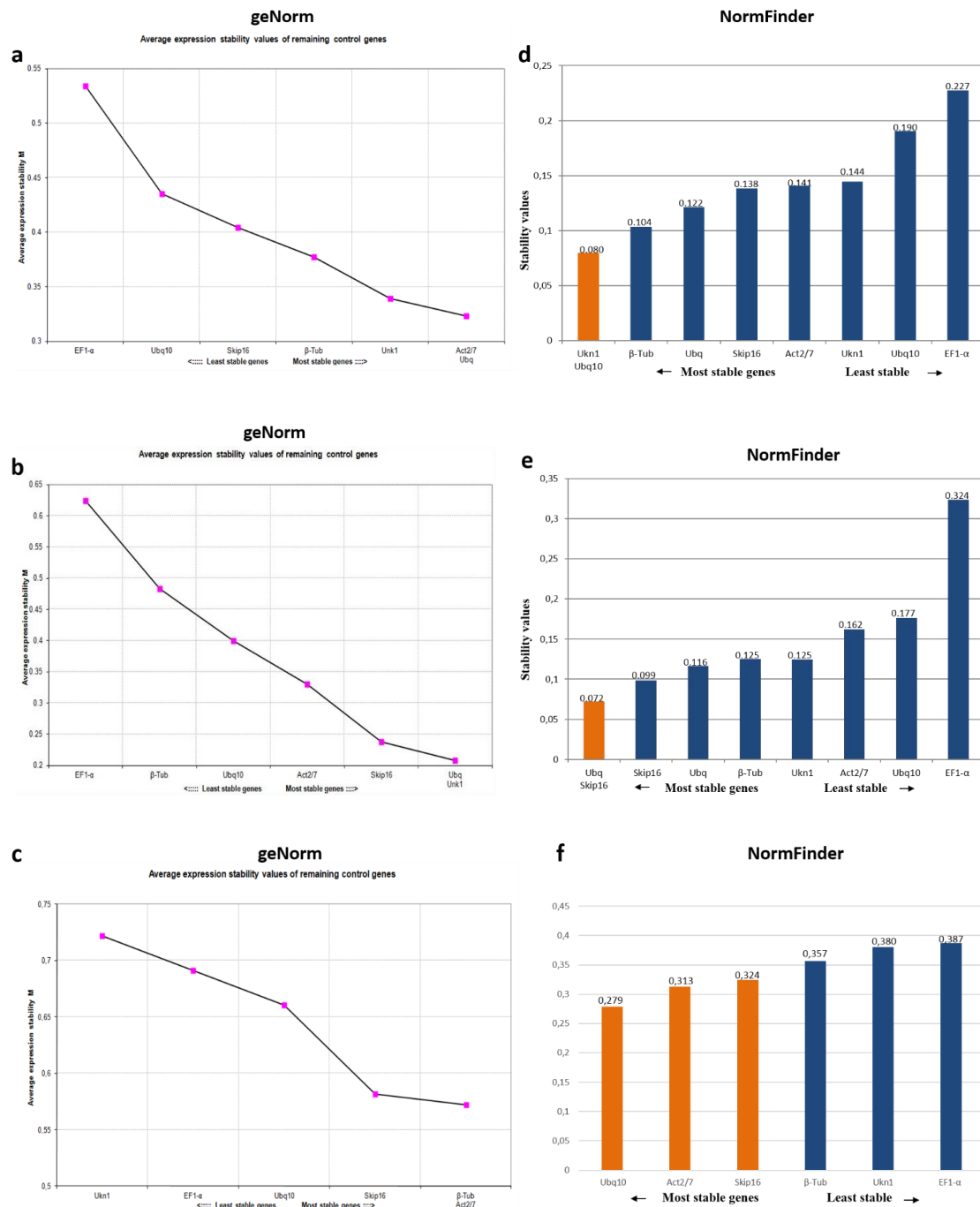


Figure 1. The expression stability of seven reference genes was evaluated by geNorm program, which calculates an average expression stability values (M) based on the average pair-wise variation existing between all pairs of candidate genes for abiotic stress and six pairs for biotic stress. The lower M value indicates a more stable expression. (a) Drought stress, (b) Salinity stress and (c) Virus stress. NormFinder ranking of expression stability of seven reference genes to abiotic stress and six genes to biotic stress according to their expression variation between inter and intra groups. The lower stability values indicate a more stable expression as a reference gene. The best combination is indicated by orange color column. (d) Drought stress, (e) Salinity stress and (f) Virus stress.

Ubiquitin is a widely used reference gene in plants with exhibits stable expression under different abiotic or biotic stresses. Our study demonstrates that *Ubq* was the most stable gene in all treatments; therefore, we propose that *Ubq* and other genes should be used as a reference gene for drought and salinity conditions in cowpea. *Ubq* had been shown to be stable under abiotic stresses in grasses, *Eucalyptus* spp., cotton and in the drought-tolerant plant *Eremosparton songoricum* [44-47], whereas it was unsuitable for normalization of common bean under abiotic stress [27] and soybean during different conditions (various developmental stages, photoperiods and cultivars) [48]. In our case, the expression of *Ubq* orthologues presented discrepancies because *Ubq10* gene was unsuitable for normalization in cowpea according to NormFinder and geNorm. This low stability was also found in previous studies in peanut and rice [49, 50]. The *Ubq* case implies that the expression stability of different reference genes from the same gene family vary in the same experimental condition. However, NormFinder chose *Ubq10* and *Ukn1* as the best combination of two genes in cowpea drought-treated roots. It is intriguing that unlike geNorm, NormFinder provides a different optimum combination in drought stress, both *Ubq10-Ukn1* which individually are only least and moderately stable, respectively, and yet, when paired, exhibit a low stability value of 0.080. In contrast, *Ukn1* was identified as being highly stable using geNorm ($M = 0.486$), which is similar to the results observed in common bean in samples subjected to different abiotic stress [27] as well as in peanut salt-treated roots [51].

Taken together, these findings indicated that each tested experimental condition demanded a specific combination of reference genes; hence, careful evaluation of the expression stability of candidate reference genes is indispensable. Several authors have used only a single gene as an internal control for normalization of RT-qPCR data. Though, normalization with two or more reference genes for RT-qPCR studies might generate more reliable results [52].

An additional important analysis performed using geNorm was the determination of optimal number of reference genes required for building an accurate normalization factor by calculating the pairwise variation (V_n/V_{n+1}) [53]. Vandesompele *et al.* [54] used 0.15 as a cut-off value, below which it was unnecessary to include an additional gene for normalization. When analyzing reference genes for our samples, we observed the $V_{2/3}$ value for drought stress

was 0.103 and V2/3 value for salinity stress was 0.077 suggesting two RGs as optimal (Figure 2). For virus stress, the pairwise variation was higher than 0.15 (0.171), so a third gene should be added.

Considering the errors that could result from single reference gene normalization strategy [54] and that at least one gene from the best couple identified by geNorm was always ranked in the third and fourth positions by NormFinder, we suggest using the specific pair of genes recommended by geNorm (Figure 1a,b,c).

As validation for our choice of genes, the best couples of RGs (according to geNorm) were chosen for normalizing the expression of three HT-SuperSAGE tags differentially expressed under salinity and drought using tolerant and sensible accessions submitted to both stresses and their negative control (Table 2). For biotic assay, only resistant accession was used.

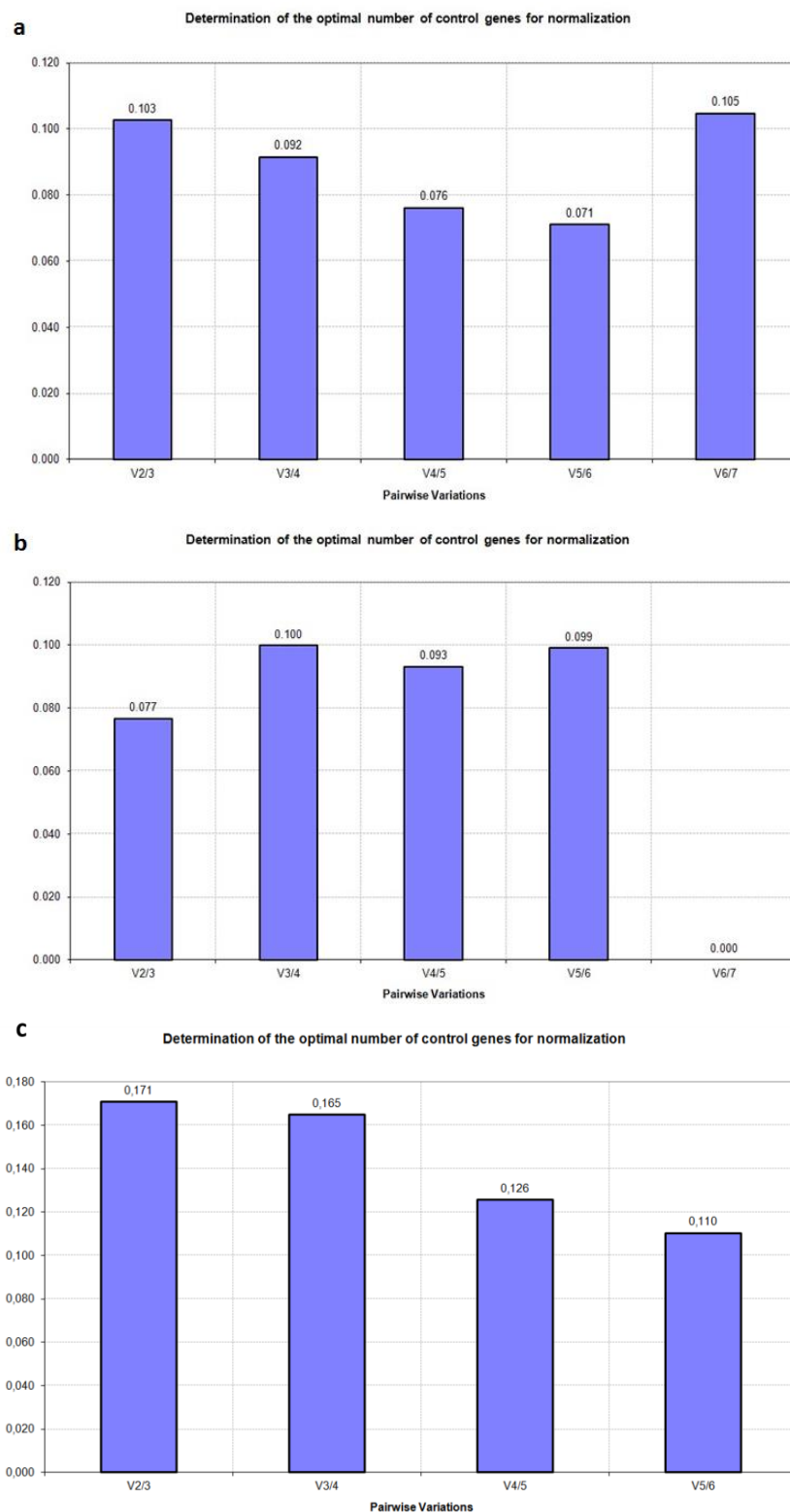


Figure 2. Evaluation of the optimum number of reference genes according to geNorm software. Pairwise variation (V) of the reference genes in (a) drought stress, (b) salinity stress, and (c) virus stress.

Table 2. Putative stress marker genes retrieved from the HT-SuperSAGE and EST analysis.

Technique and genetic material		Target		
		<i>LTP</i>	<i>CHI</i>	<i>CHS</i>
Modulation of gene expression ^{&}				
SuperSAGE* em bulks drought	Pingo de Ouro	11.21 [*] / UR	5.65 [*] / UR	47.7 [*] / UR
	Santo Inácio	-1.94 / DR	1.15 / ns	13.7 / UR
SuperSAGE* em bulks salinity	Pitiuba	2.8 / UR	-1.2 / ns	2.2 / UR
	BR14-Mulato	7.9 / UR	9.5 / UR	1.4 / ns
SuperSAGE* em bulks virus	BR14-Mulato	-3.7 / DR	5.03 / UP	-

Fold change (fc) of the most differentially expressed genes (i.e., highest and lowest fold change values) in Pingo de Ouro (drought tolerant), Santo Inácio (drought sensible), Pitiúba (salinity tolerant) and BR 14-Mulato (salinity sensible and virus tolerant) versus negative (not stressed) control obtained from the analysis of the HT-SuperSAGE. Regulation at the level of $p \leq 0.05$ - UP : up-regulated; DR: down-regulated; ns: not significant at $P < 0.05$

LTP genes are induced by drought stress and/or water deficit [55, 56]. In the HT-SuperSAGE analysis that supports this experiment, *LTP* showed different behavior in both drought contrasting accessions as shown in Table 2. The tolerant accession (Pingo de Ouro) was up-regulated and the sensitive (Santo Inácio) presented down-regulation expression. In RT-qPCR, *LTP* was up-regulated in the tolerant accession only in the third exposition time (100 min) when normalized using *Act2/7* and *Ubq*. The sensitive accession, in turn, showed no differential expression in the tree exposition times (Figure 3a). For salinity stress, the reference genes *Ubq* and *Ukn1* were used to normalize the RT-qPCR data, being induced in salinity tolerant accession (Pitiúba) in all the analyzed stress times (30, 60, 90 min. after stress), opposite to the sensitive accession (BR-14 Mulato) that did not keep this induction during the same evaluated times (Figure 3b). Thus, the results suggest that *LTP* gene may be involved in the tolerance observed in Pingo de Ouro and Pitiúba (drought and salinity tolerant, respectively), enabling the reestablishment of the proper functioning of its physiology. Our data in cowpea agree with the conclusions from the bromegrass (*Bromus inermis*) study, that during salinity stress, *LTP* mRNA accumulated within 15 min and the transcripts remained at elevated levels for the duration of the salinity stress, which suggests that this gene may play an important role in saline stress resistance [57].

LTP, involved in plant defense, was found to be down-regulated post-CPSMV infection in SuperSAGE data and RTq-PCR. These results reinforce the idea that

LTP affected the elicitor receptor involved in the control of plant defense response. An inevitable attempt was made to correlate our results with similar existing studies reporting down-regulation expression of LTPs after being infected with fungus hence allowed the pathogen to escape from host recognition [58].

Chalcone synthase (*CHS*) and Chalcone Isomerase (*CHI*) work early in the biosynthetic pathway of flavonoids, which play central roles in the interactions which occur between plants and their environment [59]. RT-qPCR analysis indicated that the *CHS* gene was expressed to a greater extent in the cowpea drought tolerant accession and in the sensitive accession at all the analyzed stress times (Figure 3a). Also, *CHS* gene is expressed at higher levels in salinity stress. The tolerant accession kept its induction over all the stress times tested and the sensitive accession showed overexpression in the two early times (Figure 3b). These results are consistent with our HT-SuperSAGE data and demonstrate that the reference genes selected from this study were reliable and form a solid base for conducting functional gene expression analysis in cowpea. These findings corroborate the results obtained by Velada and co-workers [60] who reported an increase in the *CHS* transcript accumulation upon cold stress. Interestingly, we also observed an increase in the *CHS* transcript accumulation upon virus stress, besides this tag was not available in HT-SuperSAGE, it is known that *CHS* can be induced in response to biotic stress factors, thus resulting in enhanced production of secondary metabolites [61].

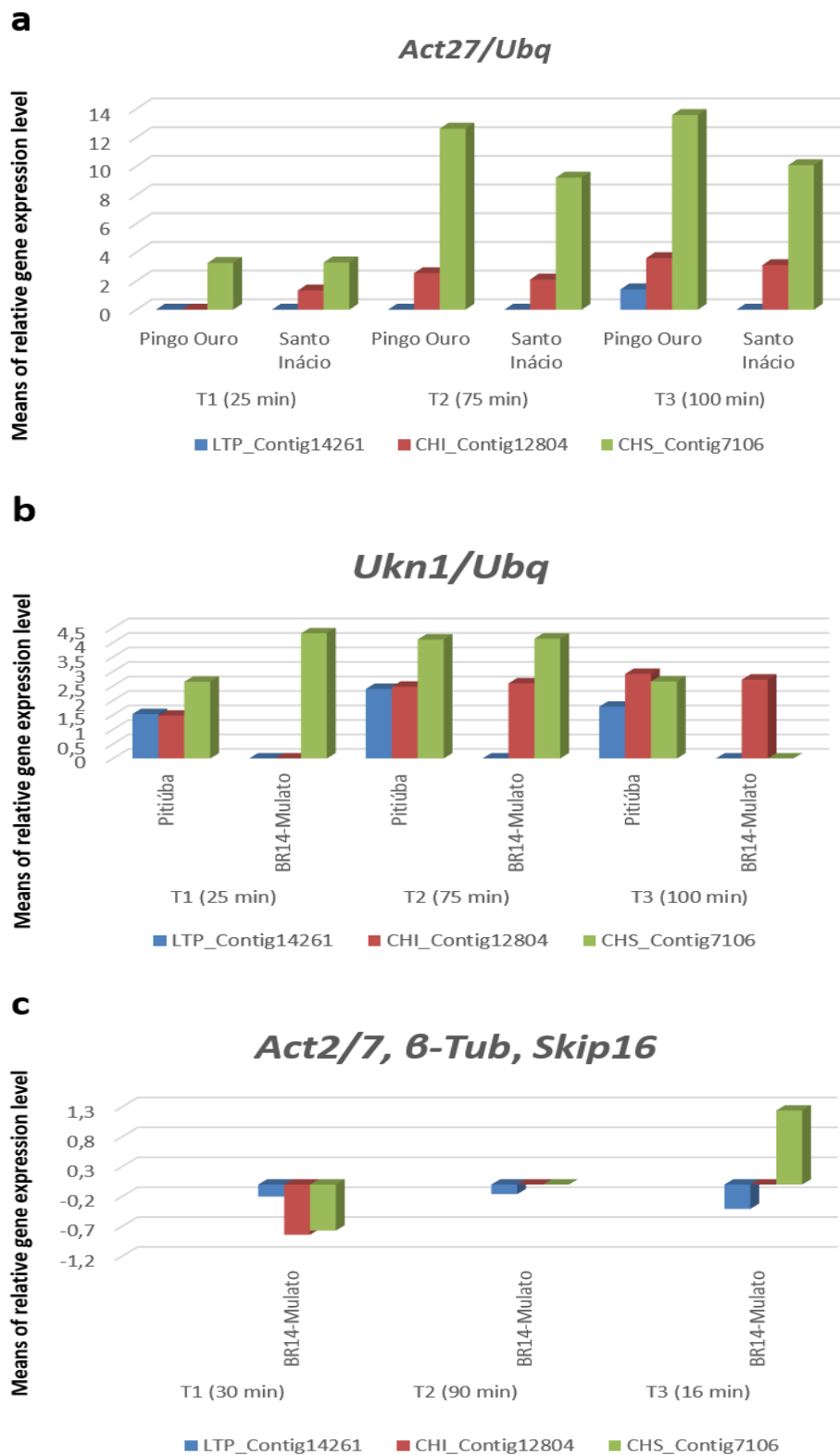


Figure 3. RT-qPCR of abiotic and biotic stress-related genes measured at the appropriate sample time using REST2009 software. The relative expression levels of genes *LTP* (blue bar), *CHI* (red bar), and *CHS* (green bar) were determined after normalization. The drought treatment samples (A) were normalized by geNorm with *Act2/7* and *Ubq*. The salinity treatment samples (B) were normalized with *Ukn1* and *Ubq*.

Moreover, *RT-qPCR* confirmed the *up-regulation* of *CHI* gene in drought tolerant cowpea at 75 and 100 min. However, *contrary to* results from HT-SuperSAGE, RTq-PCR data show that the sensitive accession was up-regulated in all the analyzed stress times. *Salt stress* increases the *expression* of *CHI* in both accessions (Pitiúba and BR14-Mulato). In the context of drought and salinity stress, flavonoids protect plants and cell constituents against oxidative damage. Therefore, our results suggest the flavonoid production in these different accessions can stabilize stress-induced oxidative stress damage [62]. Interestingly, the RT-qPCR results did not show any significant increase in *CHI* expression in cowpea during CPSMV infection.

In summary, despite slight differences found in different treatments, we concluded that at least three genes – *Ubq*, *Act2/7* and *Ukn1*– can be indicated as good reference genes for normalizing cowpea under abiotic stress.

3. Experimental Section

3.1 Plant material and Stress treatment

For root dehydration treatment, cowpea accessions ‘Pingo de Ouro’ (drought-tolerant) and ‘Santo Inácio’ (drought-sensitive) were grown in a greenhouse at Embrapa-Soybean station (Londrina, Brazil) using an aerated hydroponic system in 30 L plastic containers with pH 6.6-balanced nutrient solution as described by Kulcheski *et al.* [63]. Briefly, seeds were pre-germinated on moist filter paper in the dark at $25^{\circ}\text{C} \pm 1^{\circ}\text{C}$ and in $65\% \pm 5\%$ relative humidity. Plantlets were then placed in polystyrene supports, so the roots of the seedlings were fully immersed in the nutrient solution. Each seedling tray was maintained in a greenhouse at $25^{\circ}\text{C} \pm 2^{\circ}\text{C}$ and in $60\% \pm 5\%$ relative humidity under natural daylight (photosynthetic photon flux density (PPFD) = $1.5 \times 10^3 \mu\text{moles m}^{-2} \text{s}^{-1}$, equivalent to $8.93 \times 10^4 \text{ lux}$) for 12 h/day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) [64] were submitted to different root dehydration periods, when the nutrient solution was removed from each plastic container where the roots were kept, in the tray, in the dark, without nutrient solution or water for 0 minutes (negative control) or 25 (T1), 75 (T2), 100 minutes (T3). At the end of each period, the roots of the seedlings were immediately frozen in liquid nitrogen

and stored at -80°C until RNA extraction. The experimental design was a factorial (cultivars $\times$ extent of water deficit) with three replicates. Each replicate composed of five plantlets sampled in bulk. To avoid the impact of volatile compounds, each treatment was carried out in isolated spaces presenting the same growing conditions.

For the high-salinity treatment, NaCl was added to the Hoagland's Solution to a final concentration of 100 mM. Seeds of two contrasting cultivars, named Pitiúba (tolerant) and BR14-Mulato (sensitive) were grown in pots with washed sand for approximately 21 days. Plants grown on washed sand were watered with 200 ml of 1/2 strength Hoagland's Solution supplemented with 100 mM NaCl, and the controls, only with Hoagland Solution and water. To study relative gene expression in time-course experiment, roots were collected at 0 minutes (negative control), 30 (T1), 60 (T2) and 90 minutes (T3) under 0 or 100 mM NaCl treatments. For each time, three biological replicas were submitted to stress, and other three used as controls.

3.2 RNA isolation, quality control and cDNA synthesis and HT-SuperSAGE libraries

Total RNA was extracted using root tissues and a SV total RNA Isolation System Kits (Promega, Madison, WI) following the manufacturer's instructions and stored at -80°C . In order to verify RNA integrity isolated RNA, 5 μl of the total RNA solution were loaded on 1.5% agarose gel, electrophoresed to separate the RNA, stained with Blue Green (LGC, São Paulo, Brazil), and visualized under UV light to check the size distribution of the total RNA and the integrity of ribosomal bands. RNA yield and purity was assessed spectrophotometrically by measuring OD_{260} and $\text{OD}_{260/280}$ ratio respectively in RNase-free H_2O using Nav. For all samples $\text{OD}_{260/280}$ ratio was between 2.0 and 2.2.

The total RNA (1 μg) was reversed-transcribed into cDNA using the Improm-IITM Reverse Transcriptional System (Promega) with oligo(dT) primers following the manufacture's recommendations.

The analyzed HT-SuperSAGE libraries were produced as reported by Kido *et al.* [65] for drought stress and salinity. In the case of the drought essay, four libraries were generated using roots of two contrasting cowpea genotypes: Pingo Ouro (tolerant) and Santo Inácio (sensitive). To generate the HT-SuperSAGE

libraries, roots were collected in different times after dehydration stress (25 up to 150 minutes) in the dark, with all times bulked together to compose each stressed library, as compared with libraries using root tissues from the two not stressed controls. For the composition of this salinity stressed roots of two contrasting cowpea genotypes (Pitiúba – tolerant; BR14-Mulato – sensitive) were collected in different times after salinity stress (30, 60, 90 minutes), as compared with the negative of the same genotype. The experiment for biotic stress included three libraries. For this purpose the cowpea accession BR14-Mulato (resistant to CPSMV) was used, being the libraries composed by: (1) non stressed leaves (30 and 16h); and (2) leaves inoculated with CPSMV after injury, including a bulked library of leaves collected 30, 60 and 90 min after inoculation and (3) another library of leaves collected 16 h after virus inoculation. The libraries were constructed at GenXPro GmbH (Frankfurt, Germany) essentially as described by Matsumura *et al.* [66] and were subsequently sequenced via SOLEXA (Illumina).

3.3 Candidate reference genes and stress responsive genes from cowpea

Five candidate reference genes were selected from the literature for RT-qPCR (Table 1). These include ubiquitin, β -tubulin, elongation factor 1- α , SKP1/ASK-interacting protein 16 and unknown protein. The remaining two primer pairs were designed from cowpea EST and included the following: Actin 2/7 (Contig16004) and Polyubiquitin 10 (Contig282). The sequences of these genes were identified by BLASTn search in NordEST cowpea EST database using predicted *Arabidopsis* genes as queries. Primer pairs were designed using Primer3 web tool with the following parameter settings: the length of the primers was set to be between 20 and 22 bp, with a GC content ranging from 40% to 60% and a melting temperature (T_m value) between 57°C and 63°C (Table 3).

To validate the selected reference gene, the expression profile of 3 target genes was analyzed in the present work. The target genes, selected from the NordEST database based on their reaction to abiotic stress response and biotic stress, were *LTP*, *CHS* and *CHI*.

Table 3. Primer sequences, PCR amplicon length of the selected reference genes and stress responsive genes in cowpea.

Name	Gene ID	Cellular function	Primer Sequences	Length
<i>Act2/7</i>	AY063980 ^a	Structural constituent of cytoskeleton	F:TCAGGTGTCCAGAGGTGTTGTA R: ATGTTGTGCCTCCTGAAAGTA	120 pb
<i>β-Tub</i>	CV530631 ^a	protein binding, structural molecule activity	F: CCGTTGTGGAGCCTTACAAT R: GCTTGAGGGTCTGAAACAA	117 pb
<i>EF1-α</i>	CV530481 ^a	Translation factor activity, nucleic acid binding, GTP binding, GTPase activity	F: GGTCAATTGGTCATGTCGACTCTG R:GCACCCAGGCATACCTGAATGACC	150 pb
<i>Skip16</i>	FG231556 ^a	protein binding	F: CACCAGGATGCAAAAGTGG R: ATCCGCTTGTCCCTTGAAC	163 pb
<i>Ubq</i>	FE899660 ^a	Signaling complexes for protein degradation, translation control	F: AGAAAAGCCCCCAAGTGTTTC R: CTGCCATCTCCTTCTTCAGC	161 pb
<i>Ubq10</i>	NM_001084884 ^a		F: GTCTAAGGGGAGGAATGCAGAT R: CAAAGATCAACCTCTGCTGGTC	150 pb
<i>Ukn1</i>	AY063980 ^a	Unknown	F: ATTCCCATCATGCAGCAAAG R: AGATCCCTCCAGGTCAATCC	192 pb
<i>LTP</i>	Contig14261 _b	antimicrobial defense, signaling, and cell wall loosening	F: TGTGATGATGGAAGCGAATG R: TGAGCAGCAATCAGAGGTTG	124 pb
<i>CHI</i>	Contig12804 _b		F: CACATACCATTTCCAGCAG R: TGGAAGACACTGCCCTTGAG	149 pb
<i>CHS</i>	Contig7106 _b		F: GACTGCACAGACCATTGCAC R: GGATCGAAGGCTTCAGAAAG	144 pb

^aGenBank accession number; ^baccession number in *NordEST* genome database; F, forward primer; R, reverse primer

3.4 RT-qPCR analysis

RT-qPCR reaction was carried out in a total volume of 11.5 µl containing 7 µl of HotStart-IT SYBR green qPCR Master Mix (Usb, Affymetrix), 1 µl of 5 pM of each primer, 2 µl of template (10 X diluted cDNA from samples), 0.5 µl of ROX and 1 µl of sterile distilled water. The reaction were denatured by 95°C for 2 min and amplified by 40 cycles at 95°C for 15 s, 58°C for 30 s, and 72°C for 15 s in a LineGene 9600 Real-time PCR system (Bioer Technology Co, Bingjiang, China). The reactions were performed in triplicate, and for each biological replicate, technical triplicates were performed. Control reactions, which contained no DNA template, were conducted for all of the experiments (NTC – Non Template Control). PCR efficiency (E) of candidate genes was determined by generating standard curves. Triplicate RT-qPCR assays were applied based on a 10-fold dilution series of sample cDNA (bulk cDNA of analyzed samples). The formula $E = [10^{-1/\text{slope}}] - 1$ was applied to calculate the reaction efficiency. To pass validation, each primer must demonstrate high amplification efficiency above 80

%. All relative quantification was assessed using REST[®] v.2.0.13 software [67] REST Standard, using the pair-wise fixed randomization test with 2,000 permutations.

3.5 Statistical analyses of gene expression stability

The expression stabilities of each reference gene was calculated by geNorm version 3.5 [54] and NormFinder version 0.953 [18], which were used as described in their manuals.

4. Conclusions

We identified several reference genes that are suitable for RT-qPCR data normalization in cowpea. This is the first systematic exploration of valid reference genes in this plant species. Although most of the selected candidate genes exhibited stable expression patterns acceptable for reference genes, three evaluated genes – *Ubq*, *Act2/7* and *Ukn1* – can be indicated as the most indicated for normalizing cowpea under abiotic stress and possibly also under other experimental conditions.

Acknowledgments

This work was funded by Brazilian institutions: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Pessoal de Ensino Superior (CAPES) and Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References and Notes

1. Krasensky, J.; Jonak, C. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* **2012**, *63*, 1593–1608.
2. Chauhan, H.; Khurana, N.; Agarwal, P.; Khurana, J.P.; Khurana, P. A seed preferential heat shock transcription factor from wheat provides abiotic stress tolerance and yield enhancement in transgenic *Arabidopsis* under heat stress environment. *Plos one* **2013**, *8*, e79577.
3. Péres-Torrez, E.; Paredes C, M.; Polanco, V.; Becerra B, V. Gene Expression Analysis: A Way to Study Tolerance to Abiotic Stresses in Crops Species. *Chilean J. Agric. Res.* **2009**, *69*, 260-269.
4. Sreenivasulu, N.; Sopory, S.K.; Kavi-Kishor, P.B. Deciphering the regulatory mechanisms of abiotic tolerance in plants by genomic approaches. *Gene* **2007**, *388*, 1-13.
5. Varshney, R.K.; Close, T.J.; Singh, N.K.; Hoisington, D.A.; Cook, D.R. Orphan legumes crops enter the genomics era! *Curr. Opin. Plant Biol.* **2009**, *12*, 202-210.
6. Pottorff, M.; Wanamaker, S.; Ma, Y.Q.; Ehlers, J.D.; Roberts, P.A.; Close, T.J. Genetic and physical mapping of candidate genes for resistance to *Fusarium oxysporum* f.sp. *tracheiphilum* race 3 in cowpea [*Vigna unguiculata* L. Walp.]. *Plos one* **2012**, *7*, e41600.
7. Timko, M.P.; Singh, B.B. Cowpea, a Multifunctional legume. In *Genomics of Tropical Crop Plants*; Moore, P.H.; Ming, R., Eds.; Springer Science: New York, NY, USA, 2008; Volume 1, pp. 227-258.
8. Rocha, M.M.; Carvalho, K.J.M.; Freire Filho, F.R.; Lopes, A.C.A.; Gomes, R.L.F.; Sousa, I.S. Controle genético do comprimento do pendúculo em feijão-caupi. *Pesq. Agrop. Bras.* **2009**, *44*, 270-275.
9. Ehlers, J.D.; Hall, A.E. Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Res.* **1997**, *53*, 187-204.
10. Barrera-Figueroa, B.E.; Gao, L.; Diop, N.N.; Wu, Z.; Ehlers, J.D.; Roberts, P.A.; Close, T.J.; Zhu, J-K; Liu, R. Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. *BMC Plant Biol.* **2011**, *11*, 127.
11. Marum, L.; Miguel, A.; Ricardo, C.P.; Miguel, C. Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*. *Plos one* **2012**, *7*, e35113.
12. Wan, H.; Zhao, Z.; Qian, C.; Sui, Y.; Malik, A.A.; Chen, J. Selection of appropriate reference genes expression studies by quantitative real-time polymerase chain reaction in cucumber. *Anal. Biochem.* **2009**, *399*, 257-261.
13. Expósito-Rodríguez, M.; Borges, A.A.; Borges-Pérez, A.; Pérez, J.A. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol.* **2008**, *8*, 131.
14. Goulao, L.F.; Fortunato, A.S.; Ramalho, J.C. Selection of reference gene for normalizing quantitative real-time PCR gene expression data with multiple variables in *Coffea* spp. *Plant Mol. Biol. Rep.* **2012**, *30*, 741-759.
15. González-Agüero, M.; García-Rojas, M.; Genova, A.D.; Correa, J.; Maass, A.; Orellana, A.; Hinrichsen, P. Identification of two putative reference genes from

- grapevine suitable for gene expression analysis in berry and related tissues derived from RNA-seq data. *BMC Genomics* **2013**, *14*, 878.
16. Patel, R.K.; Jain, M. PlanRGS: A web server for the identification of most suitable candidate reference genes for quantitative gene expression studies in plant. *DNA Res.* **2011**, *18*, 463-470.
 17. Pffaffl, M.W.; Tichopad, A.; Prgomet, C.; Neuvians, T.P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlation. *Biotechnol. Lett.* **2004**, *26*, 509-515.
 18. Andersen, C.L.; Jensen, J.L.; Ørntoft, T.F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer datasets. *Cancer Res.* **2004**, *64*, 5245-5250.
 19. Vandesopelem, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**, *3*, 0034.1-0034.11.
 20. Coetzer, N.; Gazendam, I.; Oelofse, D.; Berger, D.K. SSHscreen and SSHdb, genetic software for microarray based gene discovery: application to the stress response in cowpea. *Plant Methods* **2010**, *6*, 10.
 21. Contour-Ansel, D.; Torres-Franklin, M.L.; Carvalho, M.H.C.; D'arcy-Lameta, A.; Zuily-Fodil, Y. Glutathione reductase in leaves of cowpea: cloning of two cDNAs, expression and enzymatic activity under progressive drought stress, desiccation and abscisic acid treatment. *Ann. Bot.* **2006**, *98*, 1279-1287.
 22. Huang, K.; Mellor, K.E.; Paul, S.N.; Lawson, M.J.; Mackey, A.J.; Timko, M.P. Global changes in gene expression during compatible and incompatible interactions of cowpea (*Vigna unguiculata* L.) with the root parasitic angiosperm *Striga gesnerioides*. *BMC Genomics* **2012**, *13*, 402.
 23. Pettengill, E.A.; Parmentier-Line, C.; Coleman, G.D. Evaluation of qPCR reference gene in two genotypes of *Populus* for use in photoperiod and low-temperature studies. *BMC Res. Notes* **2012**, *5*, 366.
 24. Condori, J.; Nopo-Olazabal, C.; Medrano, G.; Medina-Bolivar, F. Selection of reference gene for qPCR in hair root cultures of peanut. *BMC Res. Notes* **2012**, *4*, 392.
 25. Eitcha, D.; Zahn, M.; Bremer, M.; Yang, Z.; Rangel, A.F.; Rao, I.M.; Horst, W. Transcriptomic analysis reveals differential gene expression in response to aluminum in common bean (*Phaseolus vulgaris*) genotypes. *Ann. Bot.* **2010**, *105*, 1119-1128.
 26. Guerrero-González, S.; Rodríguez-Kessler, M.; Rodríguez-Guerra, R.; González-Chavira, M.; Simpson, J.; Sanchez, F.; Jiménez-Chavira, J.F. Differential expression of *Phaseolus vulgaris* genes induced during the interaction with *Rhizoctonia solani*. *Plant Cell Rep.* **2011**, *30*, 1465-1473.
 27. Borges, A.; Tsai, S.M.; Caldas, D.G.G. Validation of reference gene for RT-qPCR normalization in common bean during biotic and abiotic stresses. *Plant Cell Rep.* **2012**, *31*, 827-838.
 28. Vorachek, W.; Hujeriletu; Bobe, G.; Hall, J.A. Reference gene selection for quantitative PCR studies in sheep neutrophils. *Int. J. Mol. Sci.* **2013**, *14*, 11484-11495.

29. Ma, S.; Niu, H.; Liu, C.; Zhang, J.; Hou, C.; Wang, D. Expression stabilities of candidate reference genes for RT-qPCR under different stress conditions in soybean. *Plos one* **2013**, *8*, e75271.
30. Cassan-Wang, H.; Soler, M.; Yu, H.; Camargo, E.L.O.; Carocha, V.; Ladouce, N.; Savelli, B.; Paiva, J.A.P.; Leplé, J.C.; Grima-Pettenati, J. Reference genes for high-throughput quantitative reverse transcription – PCR analysis of gene expression in organs and tissues of *Eucalyptus* grown in various environmental conditions. *Plant Cell Physiol.* **2012**, *53*, 2101-2116.
31. Liu, D.S.; Shi, L.D.; Han, C.G.; Yu, J.L.; Li, D.W.; Zhang, Y.L. Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. *Plos one* **2012**, *7*, e46451.
32. Bin, W.S.; Wei, L.K.; Ping, D.W.; Li, Z.; Wei, G.; Bing, L.J.; Gui, P.B. Evaluation of appropriate reference genes for gene expression studies in pepper by quantitative real-time PCR. *Mol. Breeding* **2012**, *30*, 1393-1400.
33. Suhandono, S.; Apriyanto, A.; Ihsani, N. Isolation and characterization of three cassava elongation factor 1 alpha (MeEF1A) promoters. *Plos one* **2014**, *9*, e84692.
34. Kundun, A.; Patel, A.; Pal, A. Defining reference genes for qPCR normalization to study biotic and abiotic stress responses in *Vigna mungo*. *Plant Cell Rep.* **2013**, *32*, 1647-1658.
35. de Carvalho, K.; Filho JC, B.; dos Santos, S.G.; de Souza, S.G.; Vieira, L.G.; Pereira, L.F.; Domingues, D.S. Nitrogen starvation, salt and heat stress in coffee (*Coffea Arabica* L.) identification and validation of new genes for qPCR normalization. *Mol. Biotechnol.* **2013**, *53*, 315-325.
36. Nicot, N.; Hausman, J-F.; Hoffmann, L. Evers, D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J. Exp. Bot.* **2005**, *56*, 2907-2914.
37. Podevin, N.; Krauss, A.; Henry, I.; Swennen, R.; Remy, S. Selection and validation of reference genes for quantitative RT-PCR expression studies of the non-model crop *Musa*. *Mol. Breeding* **2012**, *30*, 1237-1252.
38. Zhu, J.; Zhang, L.; Li, W.; Han, S.; Yang, W.; Qi, L. Reference gene selection for quantitative real-time PCR normalization in *Caragana intermedia* under different abiotic stress conditions. *Plos One* **2013**, *8*, e53196.
39. Kumar, R.; Sharma, R.; Trivedi, P.C.; Vyas, G.K.; Khandelwal, V. Traditional and novel references towards systematic normalization of RT-qPCR data in plants. *AJCS* **2011**, *5*, 1455-1468.
40. Artico, S.; Nardeli, S.M.; Brilhante, O.; Grossi-de-Sa, M.F.; Alves-Ferreira, M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biol.* **2010**, *10*, 49.
41. Caldana, C.; Scheible, W.; Mueller-Roeber, B.; Ruzicic, S. A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors. *Plant Method* **2007**, *3*, 7.
42. Morgante, C.V.; Guimarães, P.M.; Martins, A.C.Q.; Araújo, A.C.G.; Leal-Bertioli, S.C.M.; Bertioli, D.J.; Brasileiro, A.C.M. Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. *BMC Res. Notes* **2011**, *4*, 339.
43. Wan, H.; Zhao, Z.; Qian, C.; Sui, Y.; Malik, A.A.; Chen, J. Selection of appropriate reference genes for gene expression studies by quantitative real-

- time polymerase chain reaction in cucumber. *Anal.Biochem* **2010**, 399, 257-261.
44. Petit, C.; Pernin, F.; Heydel, J-M.; Délye, C. Validation of a set of reference genes to study response to herbicide stress in grasses. *BMC Res. Notes* **2012**, 5, 18.
 45. Moura, J.C.M.S.; Araújo, P.; Brito, M.S.; Souza, U.R.; Viana, J.O.F.; Mazzafera, P. Validation of a set of reference genes from *Eucalyptus* spp. Under different stress conditions. *BMC Res. Notes* **2012**, 5, 634.
 46. Wang, M.; Wang, Q.; Zhang, B. Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). *Gene* **2013**, 530, 44-50.
 47. Xiao-Shuang, L.; Hong-Lan, Y.; Zhang, D-Y.; Zhang, Y-M.; Wood, A.J. Reference gene selection in the desert plant *Eremosparton songoricum*. *Int.J. Mol. Sci.* **2012**, 13, 6944-6963.
 48. Jian, B.; Liu, B.; Bi, Y.; Hou, W.; Wu, C.; Hant, T. Validation of internal control for genes expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.* **2008**, 9, 59.
 49. Yael, B.; Hovav, R. Identification of suitable internal control gene for quantitative real-time PCR expression analyses in peanut (*Arachis hypogaea*). *Peanut Sci.* **2010**, 37, 12-19.
 50. Jain, M.; Nijhawan, A.; Tyagi, A.K.; Khurana, J.P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Bioph. Res. Commun.* **2006**, 345, 646-651.
 51. Chi, X.; Hi, R.; Yang, Q.; Zhang, X.; Pan, L.; Chen, N.; Chen, M.; Yang, Z.; Wang, T.; He, Y.; Yu, S. Validation of reference gene for gene expression studies in peanut by quantitative real-time RT-PCR. *Mol.Genet. Genomics* **2012**, 287, 167-176.
 52. Gu, C.; Chen, S.; Liu, Z.; Shan, H.; Luo, H.; Guan, Z.H.; Chen, F. Reference gene selection for quantitative real-time PCR in chrysanthemum subject to biotic and abiotic stress. *Mol. Biotechnol.* **2011**, 49, 192-197.
 53. Warzybok, A.; Migocka, M. Reliable reference genes for normalization of gene expression in cumcuber grown under different nitrogen nutrition. *Plos One* **2013**, 8, e72887.
 54. Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**, 3, research0034.
 55. Colmenero-Flores, J.M.; Campos, F.; Garcarrubio, A.; Covarrubias, A.A. Characterization of *Phaseolus vulgaris* cDNA clones responsive to water deficit: identification of a novel late embryogenesis abundant-like protein. *Plant Mol. Biol.* **1997**, 35, 393-405.
 56. Jung, H.W.; Kim, W.; Hwang, B.K. Three pathogen-inducible genes encoding lipid transfer protein from pepper are differentially activated by pathogens, abiotic, and environmental stresses. *Plant Cell Environ.* **2003**, 26, 915-928.
 57. Wu, G.; Robertson, A.J.; Liu, X.; Zheng, P.; Wilen, R.W.; Nesbitt, N.T. Gusta, L.V. A lipid transfer protein gene BG-14 is differentially regulated by abiotic stress, ABA, anisomycin, and sphingosine in brome grass (*Bromus inermis*). *J. Plant Physiol.* **2004**, 161, 449-458.

58. Schlink, K. Down-regulation of defense genes and resource allocation into infected roots as factors for compatibility between *Fagus sylvatica* and. *Funct. Integr. Genomics* **2010**, *10*, 253-264.
59. Contessotto, M.G.G.; Monteiro-Vitorello, C.B.; Mariani, P.D.S.C.; Coutinho, L.L. A new member of the chalcone synthase (CHS) family in sugarcane. *Genet. Mol. Biol.* **2001**, *24*, 257–261.
60. Velada, I.; Ragonezi, C.; Arnholdt-Shmitt, B.; Cardoso, H. Reference genes selection and normalization of oxidative stress responsive genes upon different temperature stress conditions in *Hypericum perforatum* L. *PLoS ONE* **2014**, *9*(12), e115206.
61. Dao, T.T.H.; Linthorst, H.J.M.; Verpoorte, R. Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* **2011**, *10*, 397–412.
62. Pourcel, L.; Routaboul, J.M.; Cheynier, V.; Lepiniec, L.; Debeaujon, I. Flavonoid oxidation in plants: from biochemical properties to physiological functions *Trends Plant Sci.* **2007**, *12*, 29–36.
63. Kulcheski, F.R.; Oliveira, L.F.; Molina, L.G.; Almerão, M.P.; Rodrigues, F.A.; Marcolino, J.; Barbosa, J.F.; Stolf-Moreira, R.; Nepomuceno, A.L.; Marcelino-Guimarães, F.C.; *et al.* Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* **2010**, *12*, 307-324.
64. Fehr, W.R.; Caviness, C.E. *Stages of soybean development*; Ames: Iowa State University of Science and Technology, 1992. 11p.
65. Kido, E.A.; Ferreira-Neto, J.C.; Silva, R.L.O.; Belarmino, L.C.; Bezerra-Neto, J.P.; Soares-Cavalcanti, N.M.; Pandolfi, V.; Silva, M.D.; Nepomuceno, A.L.; Benko-Iseppon, A.M. Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. *BMC Bioinformatics* **2013**, *14*, S7.
66. Matsumura, H.; Molina, C.; Krüger, D.H.; Terauchi, R. Kahl G DeepSuperSAGE: High-throughput transcriptome sequencing with now- and next-Generation sequencing technologies, in tag-based next generation sequencing. In *Tag-Based Next Generation Sequencing*; Harbers, M.; Kahl, G., Eds.; Germany: Wiley-VCH Verlag GmbH & Co., 2012; pp. 3-21.
67. Pfaffl, M.W.; Horgan, G.W.; Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **2002**, *30*i e36.

6. Discussão geral

Os TFs MYB figuram como uma das maiores famílias TFs em vegetais, indicando grande potencial em conter candidatos que atuem no controle de processos fisiológicos importantes nas plantas. Nossos resultados mostraram a grande diversidade dessa família de TFs no transcriptoma do feijão-caupi, suas relações estruturais, evolutivas e sua regulação frente à infecção do CPSMV.

A ligação ao DNA pela maioria das proteínas MYB envolve a dimerização de duas hélices de reconhecimento que são hermeticamente regulados juntos. Em concordância com isso, já foi verificado que os vegetais apresentam um alto nível de conservação no domínio MYB, principalmente na segunda e terceira HTH.

O *pipeline* computacional utilizado neste estudo permitiu identificar uma quantidade de TF MYB coerente com os reportados em outras plantas. Consistente com os dados da literatura, a subfamília R2R3-MYB mostrou-se como a mais abundante entre os TFs MYB, resultante da sua extensa amplificação precedente à separação entre mono e dicotiledôneas, o que sugere que tal expansão tenha ocorrido em resposta à seleção para a regulação dos processos relacionados com o estilo de vida sésstil de plantas terrestres.

A duplicação gênica tem sido reconhecida ocorrendo ao longo da evolução das plantas, contribuindo assim para o surgimento de novas funções gênicas, subjacente às origens da novidade evolutiva. Este evento evolutivo tem sido mostrado ser significativo na expansão dos genes MYB em plantas, por meio da duplicação segmentar e duplicação em *tandem* de grandes regiões cromossômicas.

Essa grande diversidade estrutural adquirida ao longo da sua evolução resulta na funcionalidade diversificada dessa família proteica. A maior parte do conhecimento sobre os TFs MYB provém dos estudos realizados em *Arabidopsis*, os quais incluem membros atuantes no controle das respostas de defesa da planta frente ao ataque de afídeos, fungos, vírus e insetos herbívoros. A regulação positiva de três proteínas VuMYB 16 horas após a inoculação do vírus, reportado no presente trabalho, sugere que tais proteínas podem estar desempenhando papéis cruciais na regulação das vias de defesa durante os eventos de infecção na resposta de resistência nos quais são expressos.

A confiabilidade dos dados da quantificação relativa, baseada na técnica de RT-qPCR, é dependente do uso de controles endógenos, visto as inúmeras variações existentes no processo. A escolha de um *housekeeping* para uma correta normalização torna-se um pré-requisito para a confiabilidade dos resultados.

Mesmo genes envolvidos em processos celulares básicos podem ser afetados por fontes de variação experimental, assim como os genes-alvo, apresentando-se de forma mais estável em determinadas condições ambientais, implicando assim na necessidade do uso de vários genes de referência para a normalização dos resultados. Interpretações errôneas e inconsistentes podem ser passadas adiante, caso uma validação prévia para determinar os melhores genes constitutivos não seja realizada em diferentes condições experimentais.

7. Conclusões gerais

1. A cultura do feijão-caupi apresenta em seu transcriptoma, representantes das três subfamílias de TFs MYB esperadas.
2. A quantidade de genes MYB reportados para o feijão-caupi foi similar àquela descrita para outras espécies.
3. A maioria das proteínas MYB da subfamília R2R3 apresentaram um grau de conservação considerável entre os resíduos de aa em cada repetição MYB.
4. Eventos de duplicação gênica contribuíram para a expansão da família MYB no feijão-caupi.
5. Os agrupamentos formados utilizando o método de NJ permitiu inferir funções putativas às proteínas MYB do feijão-caupi, uma vez que tais proteínas compartilham estruturas dentro e fora do domínio.
6. Transcritos das subfamílias R2R3-MYB e R1-MYB estão significativamente associados com a resposta transcricional à infecção pelo CPSMV no feijão-caupi.
7. A normalização das análises de expressão gênica via RT-qPCR requer uma combinação distinta de genes endógenos de acordo com o estresse ao qual o feijão-caupi está sendo exposto.
8. As diferentes combinações entre genes da Act, Ubq, Ukn, β -tub e Skip, mostraram-se como os melhores *housekeeping* para normalização dos dados sob condições de restrição hídrica, salinidade e infecção viral.
9. O gene LTP está envolvido na resposta aos estresses hídrico e salino nas cultivares tolerantes a tais estresses. Sob infecção viral, este gene é reprimido dificultando o reconhecimento pelo hospedeiro.

8. Referências Bibliográficas

- Abe H, Urao T, Ito T, Seki M, Shinozaki K and Yamaguchi-Shinozaki K (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 15:63–78.
- Adams, MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF, *et al.* (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252(5013):1651-6.
- Alba R, Fei Z, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'ascenzo M, Gordon JS, Rose JKC, *et al.* (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* 39(5): 697-714.
- Alves MS, Dadalto SP, Gonçalves AB, de Souza GB, Barros VA and Fietto LG (2014) Transcription Factor Functional Protein-Protein Interactions in Plant Defense Responses. *Proteomes* 2: 85-106.
- Alves MS, Dadalto SP, Gonçalves AB, Souza GB, Barros VA and Fietto LG (2013) Plant bZIP Transcription Factors Responsive to Pathogens: A Review. *Int J Mol Sci* 14: 7815-7828.
- Ambawat S, Sharma P, Yadav NR and Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants* 19(3): 307–321.
- Atkinson NJ and Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. *J Exp Bot* 63(10): 3523–3544.
- Benko-Iseppon AM, Rocha MM, Freire-Filho FR, Aragão FJL, Monte SJH, Calsa-Junior T, Pandolfi V, Barbosa PK, Houllou-Kido LM, Grangeiro TB, *et al.* (2008) Genômica funcional, estrutural e comparativa de feijão-caupi (*Vigna unguiculata*). In: Anais do I Encontro de Avaliação do Programa Rede Nordeste de Biotecnologia – RENORBIO. Brasília: Ministério da Ciência e Tecnologia 1: 7-8.
- Benko-Iseppon AM, Soares-Cavalcanti NM, Wanderley-Nogueira AC, Berlarmino LC, Silva RRM, Almeida PML, Brunelli KR, Kido LMH and Kido EA (2005) Genes associated with biotic and abiotic stresses in cowpea [*Vigna unguiculata* (L.) Walp.] and other angiosperms, pp.350-359. In: Nogueira RJMC, Araújo EL, Willadino LG, Cavalcante UMT (eds) *Environmental Stresses: damages and benefits to plants*. UFRPE University Press, Recife, 500 pp.
- Benko-Iseppon, AM (2009) Identificação e Validação de Genes Importantes para a Biotecnologia e Produtividade a partir do Transcriptoma do Feijão-Caupi (*Vigna unguiculata*). Projeto Submetido ao Edital MCT/CNPq/CT-AGRO/CT-BIOTEC Nº 42/2009. Programa GENOPROT - Rede Integrada de Estudos Genômicos e Proteômicos, 34 pp.
- BIO-RAD LABORATORIES. Real-Time PCR Applications Guide. 2006. Disponível em: <<http://www.gene-quantification.de/real-time-pcr-guide-bio-rad.pdf>>.
- Brownfield L, Hafidh S, Borg M, Sidorova A and Mori T (2009) A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet* 5(3): e1000430.
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW and Shipley GL (2009) The MIQE Guidelines: Minimum

- Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* 55:4.
- Cai H, Tian S and Dong H (2012) Large Scale In Silico Identification of MYB Family Genes from Wheat Expressed Sequence Tags. *Mol Biotechnol* 52:184–192.
- Calsa Junior T, Benedito VA e Figueira AVO (2004) Análise Serial da expressão Gênica. *Revista Biotecnologia Ciência e Desenvolvimento* 33: 88-100.
- Chen X, Laudeman TW, Rushton PJ, Spraggins TA and Timko MP (2007) CGKB: An Annotation Knowledge Base for Cowpea (*Vigna unguiculata* L.) Methylation Filtered Genomic Genespace Sequences. *BMC Bioinformatics* 19(8): 129.
- Cheng H, Song S, Xiao L, Soo HM and Cheng Z (2009) Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in *Arabidopsis*. *PLoS Genet* 5(3): e1000440.
- Chi Y, Yang Y, Zhou Y, Zhou J, Fan B, Yu JQ and Chen Z (2013) Protein–Protein Interactions in the Regulation of WRKY Transcription Factors. *Mol Plant* 6(2): 287–300.
- Collinge M and Boller T (2001). Differential induction of two potato genes, *Stprx2* and *StNAC*, in response to infection by *Phytophthora infestans* and to wounding. *Plant Mol Biol* 46: 521-529.
- Companhia Nacional de Abastecimento (CONAB). 2012. Acompanhamento da Safra Brasileira de Grãos 2011/2012- Décimo Levantamento. Disponível em: http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_10_17_16_09_58_boletim_graos_-_julho_2012.pdf. Acesso em: 10/02/2013.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M (2005) Blat2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18): 3674-3676.
- Dai X, Wang Y, Yang A and Zhang WH (2012) OsMYB2P-1, an R2R3 MYB transcription factor, is involved in the regulation of phosphate-starvation responses and root architecture in rice. *Plant Physiol* 159:169–183.
- De Paula FV, Silva FF, Nascimento CS, Guimarães SEF, Martins Filho S, Nascimento M e Falcão A (2011) Métodos Estatísticos para a Avaliação da Redundância em Estudos de Etiquetas de Sequência Expressa. *Rev Bras Biom* 29(3): 462-471.
- De Vos M, Denekamp M, Dicke M, Vuylsteke M, Van Loon L, Smeeckens SC and Pieterse CM (2006) The *Arabidopsis thaliana* transcription factor AtMYB102 functions in defense against the insect herbivore *Pieris rapae*. *Plant Signal Behav* 1:305–311.
- Dias AP, Brown J, Bonello P and Grotewold E (2003) Metabolite profiling as a functional genomics tool. *Methods Mol Biol* 236: 415-26.
- Du H, Feng B, Yang S, Huang Y and Tang Y (2012) The R2R3-MYB Transcription Factor Gene Family in Maize. *BMC Plant Biol* 7: 1-12.
- Du H, Zhang L, Liu L, Tang X, Yang W, Wu Y, Huang Y, and Tang Y (2009) Biochemical and Molecular Characterization of Plant MYB Transcription Factor Family. *Biochemistry* 74(1): 1-11.
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C and Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 15(10): 573-581.
- Feller A, Machemer K, Braun EL and Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal* 66: 94-116.

- Freeman BC and Beattie GA (2008) An Overview of Plant Defenses against Pathogens and Herbivores. The Plant Health Instructor. DOI: 10.1094/PHI-I-2008-0226-01.
- Freire-Filho FR, Ribeiro VQ, Barreto PD and Santos CAF (2005) Melhoramento genético. In: Freire-Filho FR, Lima JAA and Ribeiro VQ (eds) Feijão-caupi, avanços tecnológicos. Embrapa Informação Tecnológica: Brasília, pp 28-92.
- Freire-Filho FR, Ribeiro VQ, Rocha MM, Silva KJD, Nogueira MSR and Rodrigues EV (2011) Feijão-Caupi no Brasil. Produção, Melhoramento Genético, Avanços e Desafios. 1ª Edição, Jorimá Marques Ferreira Teresina, 84 p.
- Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF and Young ND (2005) Legumes as a Model Plant Family: Genomics for Food and Feed Report of the Cross-Legume Advances through Genomics Conference. Plant Physiol 137: 1228-1235.
- Graham PH and Vance CP (2003) Legumes: Importance and Constraints to Greater Use. Plant Physiol 131: 872-877.
- Grusak MA (2002) Enhancing mineral content in plant food products. J Am Coll Nutr 21: 178S-183S.
- HarvEST (2014) (<http://harvest.ucr.edu/>)
- Heine GF (2006) Functional Analysis of P1, a model R2R3 MYB domain transcription factor. Tese (Graduate program in Plant Cellular and Molecular Biology) – Graduate School of The Ohio State University, Ohio, USA. 155pp.
- Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J Exp Bot 62(8):2465-2483.
- Jakoby M, Weisshaar MJB, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T and Parcy F (2012) bZIP transcription factors in *Arabidopsis*. Trends Plant Sci 7: 106-111.
- Jiang C, Gu X and Peterson T (2004) Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. indica. Genome Biol 5: R46.
- Jin H, Cominelli E, Bailey P, Parr A and Mehrtens F (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. EMBO J 19: 6150-6161.
- Judd WS, Campbell CS, Kellogg EA, Stevens PF and Donoghue MJ (2002) Plant Systematics: a phylogenetic approach. 2nd edition. USA, 576 pp.
- Kamiya T, Kawabe A and Miyashita NT (2002) Nucleotide polymorphism at Atmyb2 locus of wild plant *Arabidopsis thaliana*. Genet Res Camb 80:89-98.
- Klempnauer K, Gonda T and Bishop J (1982) Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. Cell 31: 453-463.
- Krishania S, Dwivedi P and Agarwal K (2013) Strategies of adaptation and injury exhibited by plants under a variety of external conditions: a short review. Com Sci 4(2): 103-110
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921.
- Lee MW, Qi M and Yang Y (2001) A novel jasmonic acid-inducible rice MYB gene associates with fungal infection and host cell death. Mol Plant Microbe Interact 14:527-535.

- Lewis, GP, Schrire, B Mackinder, B and Lock M (2005) Legumes of the world. Kew Publishing. 592pp.
- Li L, Yu X, Thompson A, Guo M, Yoshida S, Asami T, Chory J and Yin Y (2009) *Arabidopsis* MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. *Plant J* 58:275–286.
- Liao Y, Zou HF, Wang HW, Zhang WK, Ma B and Zhang JS (2008) Soybean GmMYB76, GmMYB92, and GmMYB177 genes confer stress tolerance in transgenic *Arabidopsis* plants. *Cell Res* 18:1047–1060.
- Lima HC, Queirz LP, Amorim MP, Souza VC, Dutra VF, Bortolozzu RLC, Iganci JRV, Fortunato RH, Vaz AMSF, Souza ER, *et al.* (2012) Fabaceae *In: Forzza RC et al.* (orgs) Lista de espécies da flora do Brasil. Jardim Botânico do Rio de Janeiro. (<http://floradobrasil.jbrj.gov.br/2012/FB000115>).
- Lima JAA, Sittolin IM e Lima RCA (2005) Diagnose e estratégias de controle de doenças ocasionadas por vírus. *In: Freire Filho FR, Lima JAA, RIBEIRO VQ* (Ed.) Feijão-Caupi Avanços Tecnológicos. Brasília: EMBRAPA Informação Tecnológica, pp 403-459.
- Lima LCP, Garcia FCP e Sartori ALB (2007) Leguminosae nas Florestas Estacionais do Parque Estadual do Itacolomi, Minas Gerais, Brasil: Ervas, Arbustos, Subarbustos, Lianas e Trepadeiras. *Rodriguésia* 58(2): 331-358.
- Lipsick JS (1996). One billion years of Myb. *Oncogene* 13(2): 223-35.
- Liu L, White MJ and MacRae TH (1999) Transcription factors and their genes in higher plants: Functional domains, evolution and regulation. *Eur J Biochem* 262: 247-257.
- Liu R, Lü B, Wang X, Zhang C, Zhang S, Qian J, Chen L, Shi H and Dong H (2010) Thirty-seven transcription factor genes differentially respond to a hairpin protein and affect resistance to the green peach aphid in *Arabidopsis*. *J Biosci* 35:435–450.
- Lorkowski S and Cullen P (2003) Analysing Gene Expression: A Handbook of Methods, Possibilities and Pitfalls. (Eds.) Wiley-VCH Verlag, GmbH & Co, Kga, Weinheim, Germany, Vols. 1–2, 950 pp.
- Matsumura H, Krüger DH, Kahl G and Terauchi R (2008) SuperSAGE: A modern platform for genome-wide quantitative transcript profiling. *Curr Pharm Biotechnol* 9: 368–374.
- Matsumura H, Molina C, Krüger DH, Terauchi R, Kahl G. (2012) DeepSuperSAGE: Highthroughput transcriptome sequencing with now- and next-Generation sequencing technologies. *In: Habers M, Kahl G* (eds). Tag-based next generation sequencing. 1st edition. Wiley-Blackwell, pp 3-19.
- Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, Kahl G, Reuter M, Kruger DH and Terauchi R (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc Natl Acad Sci* 100: 15718-23.
- Matsumura H, Reich S, Reuter M, Kruger DH, Winter P, Kahl G and Terauchi R (2004) SuperSAGE: A Potent Transcriptome Tool for Eukaryotic Organisms Tool for Eukaryotic Organisms. *In: San Ming Wang* (eds) Current Technologies and Applications. Horizon Scientific Press, Norwich, United Kingdom, pp 77-90.
- Matsumura H, Yoshida K, Luo S, Kimura E and Fujibe T (2010) High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PLoS One* 5: e12010.

- Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* 5: 621-628.
- Murray F, Kalla R, Jacobsen J and Gubler F (2003) A role for HvGMYB in anther development. *Plant J* 33:481–491.
- Nagaraj SH, Gasser RB and Ranganathan S (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* 8(1): 6-21.
- Nakashima K, Tran LS, Van Nguyen D, Fujita M, Maruyama K, Todaka D, Ito Y, Hayashi N, Shinozaki K and Yamaguchi-Shinozaki K (2007) Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. *The Plant J* 51: 617-630.
- Noda KI, Glover BJ, Linstead P and Martin C (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature* 369:661–664.
- Oh J E, Kwon Y, Kim JH, Noh H, Hong SW and Lee H (2011) A dual role for MYB60 in stomatal regulation and root growth of *Arabidopsis thaliana* under drought stress. *Plant Mol Biol* 77: 91-103.
- Oh SK, Lee S, Yu SH and Choi D (2005) Expression of a novel NAC domain-containing transcription factor (CaNAC1) is preferentially associated with incompatible interactions between chili pepper and pathogens. *Planta* 222: 876-887.
- Oppenheimer DG, Hertman PL, Sivakumaran S and Esch J (1991) A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67:483–493.
- Oshlack A, Robinson MD and Young MD (2010) From RNA-seq reads to differential expression results. *Genome Biol*, 11:220.
- Palmieri N and Schlotterer C. (2009) Mapping Accuracy of Short Reads from Massively Parallel Sequencing and the Implications for Quantitative Expression Profiling. *PLoS One* 4(7): e6323.
- Pandey A and Mann M (2000) Proteomics to study genes and genomes. *Nature* 405: 837-846.
- Park JS, Kim JB, Cho KJ, Cheon CI, Sung MK, Choung MG and Roh KH (2008) *Arabidopsis* R2R3-MYB transcription factor AtMYB60 functions as a transcriptional repressor of anthocyanin biosynthesis in lettuce (*Lactuca sativa*). *Plant Cell Rep* 27:985–994.
- Pinto MST, Ribeiro JM e Oliveira EAG (2011) O estudo de genes e proteínas de defesa em plantas. *R Bras Bioci* 9(2): 241-248.
- Pollock JD (2002) Gene expression profiling: methodological challenges, results, and prospects for addiction research. *Chem Phys Lipids* 121 241-256.
- Raffaele S, Rivas S and Roby D (2006) An essential role for salicylic acid in AtMYB30-mediated control of the hypersensitive cell death program in *Arabidopsis*. *FEBS Lett* 580:3498–3504.
- Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW and Velculescu VE (2002) Using the transcriptome to annotate the genome. *Nat Biotechnol* 20(5): 508-12.
- Schena M, Shalon D, Davis RW and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-70.

- Seki M, Kamei A, Yamaguchi-Shinozaki K and Shinozaki K (2003) Molecular responses to drought salinity and frost: common and different paths for plant protection. *Curr Opin Biotech* 14:194-199.
- Seo PJ and Park CM (2010) MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in *Arabidopsis*. *New Phytol* 186:471-483.
- Seo PJ, Xiang F, Qiao M, Park JY, Lee YN, Kim SG, Lee YH, Park WJ and Park CM (2009) The MYB96 Transcription Factor Mediates Absciscic Acid Signaling during Drought Stress Response in *Arabidopsis*. *Plant Physiol* 151: 275-289.
- Shinkets RA. Gene expression profiling. 1^o ed. Totowa: Humana Press, Inc, 2004.
- Shin D, Moon S, Han S, Kim B, Park SR, Lee S, Yoon H, Lee HE, Kwon H, Baek D *et al.* (2011) Expression of StMYB1R-1, a novel potato single MYB-like domain transcription factor, increases drought tolerance. *Plant Physiol* 155: 421-432.
- Shipley, GL (2006) An introduction to real-time PCR. In: Dorak MT (ed) Real-Time PCR. Taylor & Francis Group, New York, pp. 1-31.
- Shivashankar HN, Gasser RB and Ranganathan S (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings Bioinf* 8(1) 6-21.
- Singh KB, Foley RC and Sánchez LO (2002) Transcription factors in plant defense and stress responses. *Curr Opin Plant Biol* 5:430-436.
- Sinicropi D, Cronin M, Liu ML (2007) Gene Expression Profiling Utilizing Microarray Technology and RT-PCR. In: Ferrari ME-i-c, Ozkan M, Heller M, editors. *BioMEMS and Biomedical Nanotechnology, Volume II: Micro/Nano Technologies for Genomics and Proteomics*, Springer, 540 pp.
- Souza NM and Souza LAG (2011) Levantamento do potencial de aproveitamento das leguminosas no distrito da Barreira do Andirá, Barreirinha, AM. *Enciclopédia Biosfera* 7 (12): 1-23.
- Souza VC and Lorenzi H (2005) Botânica sistemática: guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG II. Nova Odessa, São Paulo, 639 pp.
- Stracke R, Favory JJ, Gruber H, Bartelniewoehner L, Bartels S, Binkert M, Funk M, Weisshaar B and Ulm R (2010) The *Arabidopsis* bZIP transcription factor HY5 regulates expression of the PFG1/ MYB12 gene in response to light and ultraviolet-B radiation. *Plant Cell Environ* 33:88-103
- Stracke R, Werber M and Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* 4: 447-456.
- Terauchi R, Matsumura H, Krüger DH and Kahl G (2008) SuperSAGE: The most advanced transcriptoma technology for functional genomics. In: Kahl, Günter; Meksem, Khalid (eds.). *The handbook of plant functional genomics*. 1st edition. Wiley-VCH Verlag GmbH & Co, Weinheim, pp 37-54.
- Timko MP, Ehlers JD and Roberts PA (2007) Cowpea: pulses, sugar and tuber crops. *Genome Mapp Mol Breed Plants* 3: 49-67.
- Tombuloglu H, Kekec G, Sakcali MS and Turgay Unver (2013) Transcriptome-wide identification of R2R3-MYB transcription factors in barley with their boron responsive expression analysis. *Mol Genet Genomics* 288(3-4): 141-155.
- Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, Zhang J, Benedito V, Hofer JMI and Chueng F, *et al.* (2007) Legume transcription

- factors: global regulators of plant development and response to the environment. *Plant Phys* 144: 538-549.
- Ülker B and Somssich IE (2004) WRKY transcription factors: from DNA binding towards biological function. *Curr Opin Plant Biol* 7:491–498.
- Vance CP, Grogan PH and Allan DL (2000) Biological nitrogen fixation. Phosphorus a critical future need. *In*: Pedrosa FO, Hungria M, Yates MG and Newton WE (eds) *Nitrogen fixation: from molecules to crop productivity*. 1st edition. Kluwer Academy Publishers, Dordrecht, pp 506-514.
- Vanhauwaert S, Peer GV, Rihani A, Janssens E, Rondou P, Lefever S, Paepe A, Coucke PJ, Speleman F, Vandesompele J and Willaert A (2014) Expressed Repeat Elements Improve RT-qPCR Normalization across a Wide Range of Zebrafish Gene Expression Studies. *PLoS ONE* 9(10): e109091.
- Velázquez SF, Guerra RR and Calderón LS (2011) Abiotic and Biotic Stress Response Crosstalk in Plants. *In*: Shanker, A. (ed) *Abiotic Stress Response in Plants. Physiological, Biochemical and Genetic Perspectives*. InTech, Rijeka, Croatia, 346 pp.
- Velculescu VE, Zhang L, Vogelstein B and Kinzler KW (1995) Serial analysis of gene expression. *Science* 270: 484-487.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.* (2001) The Sequence of the Human Genome. *Science* 291:1304–1351.
- Verdier J, Zhao J, Torres-Jerez I, Ge S, Liu C, He X, Mysore KS, Dixon R and Udvardi MK (2012) MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. *PNAS* 109(5):1766–1771.
- Voitsik AM, Muench S, Deising HB and Voll LM (2013) Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. *BMC Plant Biology* 13 (85): 1-16.
- Wanderley-Nogueira AC, Soares-Cavalcanti NM, Morais DAL, Belarmino LC, Barbosa-Silva A and Benko-Iseppon AM (2007) Abundance and diversity of resistance genes in the sugarcane transcriptome revealed by in silico analysis. *Genet Mol Res* 6(4): 866-889.
- Wang SM (2007) Understanding SAGE data. *Trends Genet* 23(1): 42-50.
- Wang X, Basnayake BMVS, Zhang H, Li G, Li W, Mengiste NVT and Song F (2009a) The *Arabidopsis* ATAF1, a NAC Transcription Factor, Is a Negative Regulator of Defense Responses Against Necrotrophic Fungal and Bacterial Pathogens. *Mol Plant Microbe Interact* 22 (10): 1227–1238.
- Wang X, Niu QW, Teng C, Li C, Mu J, Chua NH and Zuo J. (2008) Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in *Arabidopsis*. *Cell Res* 19:224–235.
- Wang Z, Gerstein M and Snyder M (2009b) RNA-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1): 57–63.
- Wang Z, Libault M, Joshi T, Valliyodan B, Nguyen HT, Xu D, Stacey G, and Cheng J (2010) SoyDB: a knowledge database of soybean transcription factors. *Plant Biol* 10: 1-12.
- Weston K (1998). Myb proteins in life, death and differentiation. *Curr Opin Genet Dev* 8(1): 76-81.
- Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol Ecol Resour* 13: 559–572.

- Xie RJ, Zheng L, Deng L, He SL, Yi SL, Lv Q and Zheng YQ (2014) The Role Of R2R3MYB Transcription Factors In Plant Stress Tolerance. *The JAPS* 24(6): 1821-1833.
- Yan Y, Borrego E and Kolomiets MV (2013) Jasmonate biosynthesis, perception and function in plant development and stress responses. In: Baez RV (ed) *Lipid metabolism*. InTech Europe, Rijeka, Croatia, pp 393–442.
- Zhang Y, Cao G, Qu LJ and Gu H (2009) Characterization of *Arabidopsis* MYB transcription factor gene AtMYB17 and its possible regulation by LEAFY and AGL15. *J Genet Genomics* 36:99–107.

9. Anexos

Anexo 1: Sequências-isca (*seed sequences*) de TF MYB obtidas no banco de dados *Plant Transcription Factor Database* (<http://planttfdb.cbi.pku.edu.cn/index.php>) a partir do organismo modelo *Arabidopsis thaliana*.

TF ID	Nome Comum	Tamanho		Resposta em Processos Biológicos
		Proteína	nucleotídeo	
AT1G06180.1	ATMYB13 / ATMYBLFGN / MYB13	246 aa	912 pb	Estresse salino / Estímulo dos fitohormônios giberelina, ácido salicílico e ácido jasmônico
AT1G08810.1	AtMYB60 / MYB60	280 aa	1123 pb	Estresse hídrico / Estímulo da luz / Movimento dos estômatos / Estímulo ao ácido abscísico, ácido salicílico e ácido jasmônico
AT1G09540.1	ATMYB61 / MYB61	366 aa	1498 pb	Desenvolvimento vascular, radicular, do xilema e do revestimento da semente / Movimento dos estômatos / Biossíntese de brassinosteróides / Estímulo da auxina
AT1G17950.1	ATMYB52, BW52, MYB52	249 aa	1123 pb	Regulação de processo na biogênese da parede celular secundária / Estímulo ao ácido abscísico
AT1G18570.1	AtMYB51, BW51A, BW51B, HIG1, MYB51	352 aa	1654 pb	Cascata MAPK / Resposta de defesa contra bactérias, insetos e fungos / Estímulo à fitohormônios / Regulação da resposta imune, imune inata e de hipersensibilidade / SAR / Resposta ao estresse salino
AT1G22640.1	ATMYB3, MYB3	257 aa	1077 pb	Estresse salino / Estímulo aos fitohormônios geberelina, ácido abscísico, ácido salicílico, ácido jasmônico, etileno e auxina / Resposta a ferimentos / Regulação negativa do processo metabólico
AT1G34670.1	AtMYB93, MYB93	365 aa	1098 pb	Estímulo de auxina, ácido abscísico e ácido salicílico

Continuação anexo 1: Sequências iscas (*seed sequence*) de TF MYB obtidas no banco de dados *Plant Transcription Factor Database* (<http://planttfdb.cbi.pku.edu.cn/index.php>) a partir do organismo modelo *Arabidopsis thaliana*.

TF ID	Nome Comum	Tamanho		Resposta em Processos Biológicos
		Proteína	nucleotídeo	
AT1G49010.1	-	314 aa	1161 pb	Resposta ao estresse salino / Estímulo aos fitohormônios giberelina, auxina, ácido jasmônico e ácido salicílico / Resposta a íons cádmio / Modificação do RNAm
AT2G32460.2	ATM1, ATMYB101, MYB101	478 aa	1792 pb	Desenvolvimento do pólen / Processo biossintético e via de sinalização da giberelina / Regulação positiva da transcrição, morte celular programada e via de sinalização mediada pelo ácido abscísico
AT2G37630.1	AS1, ATMYB91, ATPHAN, MYB91	367 aa	1490 pb	Resposta de defesa contra vírus, bactérias, fungos e estresse salino / Estímulo aos fitohormônios giberelina, auxina, ácido salicílico, ácido jasmônico / regulação da resposta imune inata / regulação negativa da transcrição (DNA-dependente)
AT2G47460.1	ATMYB12, MYB12, PFG1	371 aa	1453 pb	Estímulo ao etileno, auxina e sacarose / Resposta à exposição UV-B / Biossíntese de flavonóide
AT3G11280.1	-	263 aa	1242 pb	Estímulo ao fitohormônio giberelina e ácido salicílico / Biossíntese da cisteína
AT3G11440.1	ATMYB65, MYB65	553 aa	2015 pb	Estímulo ao ácido salicílico, etileno e giberelina / Processo biossintético da giberelina / Regulação positiva de morte celular programada / Regulação negativa de crescimento
AT3G18100.1	AtMYB4R1, MYB4R1	847 aa	2655 pb	Processo biossintético do ácido abscísico
AT3G30210.1	ATMYB121, MYB121	276 aa	1224 pb	Estímulo ao ácido abscísico

Continuação anexo 1: Sequências iscas (*seed sequence*) de TF MYB obtidas no banco de dados *Plant Transcription Factor Database* (<http://planttfdb.cbi.pku.edu.cn/index.php>) a partir do organismo modelo *Arabidopsis thaliana*.

TF ID	Nome Comum	Tamanho		Resposta em Processos Biológicos
		proteína	nucleotídeo	
AT3G46130.1	ATMYB48, ATMYB48-1, ATMYB48-2, ATMYB48-3, MYB48	256 aa	771 pb	Estímulo ao ácido salicílico
AT3G61250.1	AtMYB17, LMI2, MYB17	299 aa	1391 pb	Estímulo ao ácido salicílico e ácido jasmônico / Regulação do desenvolvimento da flor e estame
AT4G09460.1	AtMYB6, MYB6	236 aa	984 pb	Estímulo a giberelina, ácido salicílico e ácido jasmônico
AT5G06100.1	ATMYB33, MYB33	451 aa	2127 pb	Via de sinalização mediada pela giberelina e ácido abscísico / Processo biossintético da giberelina / Regulação positiva da morte celular programada e negativa do crescimento
AT5G16770.1	AtMYB9, MYB9	336 aa	1284 pb	Estímulo ao ácido abscísico
AT5G15310.2	ATMIXTA, ATMYB16, MYB16	315 aa	1325 pb	Resposta ao estresse salino / Estímulo aos fitohormônios giberelina, auxina, etileno, ácido jasmônico e ácido salicílico / Resposta a íons cádmio
AT5G15310.1	ATMIXTA, ATMYB16, MYB16	326 aa	1325 pb	Resposta ao estresse salino / Estímulo aos fitohormônios giberelina, auxina, etileno, ácido jasmônico e ácido salicílico / Resposta a íons cádmio
AT1G09770.1	ATCDC5, ATMYBCDC5, CDC5	844 aa	2934 pb	Ciclo celular mitótico / Processamento do RNA / Vias de sinalização de resposta de defesa (gene-dependente e gene-independente) contra bactérias e fungos
AT3G09230.1	AtMYB1, MYB1	393 aa	1448 pb	Estímulo ao ácido salicílico

Anexo 2: Caracterização das sequências de TF da família MYB identificadas no transcriptoma do feijão-caupi.

tBlastn	Blastx										ORF finder						
Cluster	e-value	Score	Frame	Identity	positivity	gaps	length	Acesso	Organism	Description	evalue	score	Identity	Positivity	Frame	length	domain
SI22690.2	1.00E-14	80.1	-1	38/65 (58%)	47/65 (71%)		2208	ESW14862.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_007G023600g	1,00E-71	244	119/132(90%)	124/132(93%)	-1	447	1528..1974
PO12405.1	2,00E-19	95.9	2	48/115 (41%)	68/115 (58%)	1/115 (0%)	2023	AGV54357.1	<i>Phaseolus vulgaris</i>	MYB transcription factor	9,00E-106	336	164/166(99%)	165/166(99%)	2	549	215..763
SI39136.1	1,00E-43	176	1	80/116 (68%)	95/116 (80%)		2003	XP_003553691.1	<i>Glycine max</i>	PREDICTED: uncharacterized protein LOC100788976	0.0	619	373/453(82%)	392/453(86%)	1	1353	229..1581
SI28726.1	2,00E-18	92.4	-1	43/103 (41%)	62/103 (59%)	1/103 (0%)	1826	ESW33891.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_001G106800g	0.0	632	322/362(89%)	330/362(91%)	-1	1083	330..1412
SI30097.1	5,00E-18	90.9	1	42/97 (43%)	60/97 (61%)	1/97 (1%)	1762	ESW23509.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_004G053500g	2,00E-178	521	297/361(82%)	312/361(86%)	1	1068	166..1233
SI42807.1	2,00E-45	181	-2	77/112 (68%)	90/112 (79%)		1644	XP_003543981.1	<i>Glycine max</i>	PREDICTED: transcription repressor MYB5-like	7,00E-134	406	239/323(74%)	264/323(81%)	-2	1113	441..1553
SI40726.1	1,00E-38	159	-2	75/125 (60%)	93/125 (74%)	1/125 (0%)	1373	XP_003543751.1	<i>Glycine max</i>	PREDICTED: transcription factor RAX1	0.0	577	303/362(84%)	322/362(88%)	-2	1086	185..1270
SI15799.1	1,00E-38	159	-3	76/128 (59%)	94/128 (73%)	8/128 (6%)	1352	XP_003550598.1	<i>Glycine max</i>	PREDICTED: transcription factor RAX3-like	7,00E-155	454	262/340(77%)	284/340(83%)	-3	963	268..1230
SI9693.1	5,00E-43	173	1	77/125 (61%)	97/125 (77%)	1/125 (0%)	1295	XP_003535515.1	<i>Glycine max</i>	PREDICTED: transcription factor TT2-like	2,00E-151	442	222/287(77%)	238/287(82%)	1	819	46..864
SI33857.3	1,00E-18	93.2	-3	43/73 (58%)	53/73 (71%)		1133	ESW10131.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_009G183300g	1,00E-55	197	162/220(74%)	173/220(78%)	-3	498	439..936
FF545726	2,00E-048	191	1	90/133 (67%)	106/133 (79%)	2/133 (1%)	862	XP_003534572.1	<i>Glycine max</i>	PREDICTED: myb-related protein 315-like	3,00E-137	402	206/236(87%)	214/236(90%)	1	720	133..852
SI69686.1	2,00E-13	75.9	-1	41/121 (33%)	64/121 (52%)		833	XP_005779616.1	<i>Emiliania huxleyi</i>	hypothetical protein EMIHUDRAFT_42066, partial	1,00E-26	110	49/97(51%)	66/97(68%)	-1	798	12..809
VuMYB1	2,00E-15	82.4	2	44/123 (35%)	63/123 (50%)	10/123 (8%)	3594	ESW29665.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_002G088900g	0.0	832	425/440(97%)	432/440(98%)	1	1386	535..1920
VuMYB4	3,00E-35	147	-1	63/104 (60%)	81/104 (77%)		2238	ESW20765.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_005G012900g	0.0	852	430/485(89%)	453/485(93%)	1	1368	2563..3930
VuMYB5	3,00E-19	95.5	-1	43/71 (60%)	53/71 (74%)		1475	SW24911.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_004G171200g	7,00E-133	399	229/269(85%)	235/269(87%)	-1	768	494..1261
VuMYB7	7,00E-05	47.4	1	22/47 (46%)	32/47 (67%)		4141	ESW13956.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_008G240700g	0.0	728	391/450(87%)	404/450(89%)	3	1272	2286..3557
VuMYB8	1,00E-31	136	-2	79/197 (40%)	114/197 (57%)	2/197 (1%)	1397	ESW35507.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_001G240300g	0.0	523	249/299(83%)	264/299(88%)	1	849	568..1416
VuMYB9	1,00E-26	107	-2	51/89 (57%)	62/89 (69%)		2184	ESW22285.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_005G141200g	2,00E-96	304	167/205(81%)	173/205(84%)	2	525	803..1327
VuMYB10	3,00E-049	194	-2	85/116 (73%)	99/116 (85%)		1585	ESW31233.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_002G221000g	0.0	605	335/371(90%)	342/371(92%)	-1	1083	301..1383
VuMYB11	2,00E-45	181	1	80/116 (68%)	93/116 (79%)		918	ESW17115.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_007G211800g	5,00E-108	336	183/220(83%)	190/220(86%)	-1	741	643..1383
VuMYB13	3,00E-35	148	1	84/202 (41%)	119/202 (58%)	1/202 (0%)	2648	ESW13792.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_008G226600g	0.0	609	288/305(94%)	295/305(96%)	1	933	688..1620
VuMYB15	7,00E-05	47.4	3	24/45 (53%)	31/45 (68%)		1477	ESW16515.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_007G162800g	2,00E-158	469	248/276(90%)	259/276(93%)	3	873	75..947
VuMYB17	4,00E-66	251	3	147/293 (50%)	173/293 (58%)	18/293 (6%)	1298	ESW27133.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_003G176800g	5,00E-170	495	293/311(94%)	299/311(96%)	1	936	355..1290
VuMYB18	2E-19	96.3	-2	58/150 (38%)	77/150 (50%)	3/150 (2%)	1219	ESW10131.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_009G183300g	1,00E-117	360	265/344(77%)	279/344(81%)	-3	849	408..1256
VuMYB19	2E-20	99.0	-3	51/114 (44%)	66/114 (57%)	2/114 (1%)	804	ESW24609.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_004G144900g	2,00E-141	416	223/236(94%)	227/236(96%)	-3	714	295..1008
VuMYB21	3E-46	184	-2	82/116 (70%)	98/116 (83%)		1209	ESW29070.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_002G040900g	2,00E-166	481	256/323(79%)	277/323(85%)	2	912	86..997
VuMYB22	2,00E-34	145	-1	62/103 (60%)	76/103 (73%)		1000	ESW11415.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_008G028000g	5,00E-152	441	223/266(84%)	239/266(89%)	2	864	98..961
VuMYB23	4E-53	207	-3	119/269 (44%)	150/269 (55%)		2534	ESW25653.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_003G054100g	1,00E-155	469	279/294(95%)	286/294(97%)	1	885	787..1671
VuMYB24	7E-41	166	2	76/116 (65%)	92/116 (78%)		1885	ESW35249.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_001G219000g	0.0	749	380/404(94%)	385/404(95%)	3	1215	585..1799
VuMYB26	4E-45	180	2	80/116 (68%)	92/116 (78%)		1133	ESW09345.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_009G119900g	2,00E-146	432	239/259(92%)	244/259(94%)	-1	786	404..1189
VuMYB27	2E-19	95.9	2	48/115 (41%)	68/115 (58%)	1/115 (0%)	1574	AGV54357.1	<i>Phaseolus vulgaris</i>	MYB transcription factor	0.0	667	325/333(98%)	329/333(98%)	2	1002	2654..3655
VuMYB28	0,000001	53.1	-2	28/90 (31%)	47/90 (52%)		2177	ESW25069.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_003G004600g	0.0	871	492/584(84%)	512/584(87%)	3	1506	1638..3143
VuMYB29	4,00E-16	84.3	2	42/121 (34%)	64/121 (52%)	2/121 (1%)	1663	ESW05985.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_010G009800g	0.0	583	322/352(91%)	335/352(95%)	-1	1074	397..1470
VuMYB30	0,00000006	57.4	2	38/121 (31%)	63/121 (51%)	10/121 (8%)	830	NP_001236400.1	<i>Glycine max</i>	MYB transcription factor MYB114	0.0	729	514/757(68%)	573/757(75%)	3	2010	825..2834
VuMYB31	1E-47	189	2	84/117 (71%)	97/117 (82%)		1158	ESW09812.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_009G158200g	6,00E-162	489	296/322(92%)	302/322(93%)	2	969	83..1051
VuMYB32	7E-36	150	-1	62/106 (58%)	84/106 (78%)		2339	ESW18185.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_006G020200g	0.0	597	286/332(86%)	301/332(90%)	-2	1110	677..1786
VuMYB33	9E-34	143	2	83/203 (40%)	112/203 (54%)		1750	ESW12375.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_008G107000g	0.0	547	275/286(96%)	280/286(97%)	-2	861	1040..1900
VuMYB35	9E-55	213	-1	119/269 (44%)	150/269 (55%)		1685	ESW25172.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_003G013600g	0.0	546	283/296(96%)	288/296(97%)	2	891	578..1468
VuMYB37	7E-41	166	-2	78/116 (67%)	91/116 (78%)		1579	ESW10019.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_009G174900g	0.0	642	333/357(93%)	345/357(96%)	-3	1074	274..1347
VuMYB38	4E-46	184	-1	104/259 (40%)	131/259 (50%)	22/259 (8%)	1072	ESW04606.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_011G109700g	3,00E-153	451	278/327(85%)	293/327(89%)	3	960	120..1079
VuMYB39	1E-44	179	-1	82/125 (65%)	97/125 (77%)		922	ESW04605.1	<i>Phaseolus vulgaris</i>	hypothetical protein PHAVU_011G109600g	1,00E-157	460	254/323(79%)	270/323(83%)	1	906	118..1023

Continuação Anexo 2: Caracterização das sequências de TF da família MYB identificadas no transcriptoma do feijão-caupi.

tBlastn	Blastx															ORF finder			
Cluster	e-value	Score	Frame	Identity	positivity	gaps	length	Acesso	Organism	Description	evalue	score	Identity	Positivity	Frame	Frame	length	domain	
VuMYB40	9E-38	156	3	72/117 (61%)	88/117 (74%)	1/117 (0%)	1206	ESW34769.1	Phaseolus vulgaris	hypothetical protein PHAVU_001G179400g	7,00E-159	462	284/333(85%)	297/333(89%)	-1	-1	933	170..1102	
VuMYB44	0,000002	52.8	1	27/68 (39%)	41/68 (59%)		3067	ESW24591.1	Phaseolus vulgaris	hypothetical protein PHAVU_004G143400g	0.0	864	445/475(94%)	458/475(96%)	1	1	1428	262..1689	
VuMYB45	1E-19	96.7	1	45/73 (61%)	54/73 (73%)		1369	ESW29544.1	Phaseolus vulgaris	hypothetical protein PHAVU_002G078600g	4,00E-131	399	281/303(93%)	287/303(94%)	3	3	912	237..1148	
VuMYB46	2E-19	95.9	-2	43/74 (58%)	56/74 (75%)		1522	ESW27366.1	Phaseolus vulgaris	hypothetical protein PHAVU_003G195700g	1,00E-163	480	261/300(87%)	278/300(92%)	3	3	903	168..1070	
VuMYB49	4E-34	144	1	67/119 (56%)	83/119 (69%)		678	ESW05819.1	Phaseolus vulgaris	hypothetical protein PHAVU_011G212000g	8,00E-130	392	220/307(72%)	247/307(80%)	2	2	885	170..1054	
VuMYB52	8E-29	127	2	88/250 (35%)	114/250 (45%)	11/250 (4%)	649	ESW05986.1	Phaseolus vulgaris	hypothetical protein PHAVU_010G009900g	1,00E-170	496	305/350(87%)	314/350(89%)	-3	-3	1053	192..1244	
VuMYB53	2E-43	175	-3	79/109 (72%)	89/109 (81%)		1359	ESW21643.1	Phaseolus vulgaris	hypothetical protein PHAVU_005G087400g	1,00E-97	235	172/248(69%)	195/248(78%)	-1	-3	753	804..1271	
VuMYB54	7E-43	173	-1	79/120 (65%)	91/120 (75%)		1267	ESW11573.1	Phaseolus vulgaris	hypothetical protein PHAVU_008G041500g	9,00E-176	506	300/324(93%)	313/324(96%)	-1	-1	972	155..1126	
VuMYB56	3E-48	191	-1	81/126 (64%)	100/126 (79%)		1210	ESW24277.1	Phaseolus vulgaris	hypothetical protein PHAVU_004G116500g	3,00E-148	434	218/275(79%)	238/275(86%)	-1	-1	867	251..1117	
VuMYB57	1E-19	96.7	1	44/76 (57%)	55/76 (71%)		726	ESW32232.1	Phaseolus vulgaris	hypothetical protein PHAVU_002G304400g	4,00E-114	342	200/227(88%)	206/227(90%)	-3	-3	678	216..893	
VuMYB60	8E-23	106	-2	50/106 (47%)	71/106 (66%)	1/106 (0%)	3488	ESW33338.1	Phaseolus vulgaris	hypothetical protein PHAVU_001G061200g	0.0	1784	885/1011(88%)	930/1011(91%)	2	2	3012	365..3376	
VuMYB61	0.0	780	3	398/620 (64%)	479/620 (77%)	3/620 (0%)	2578	ESW23966.1	Phaseolus vulgaris	hypothetical protein PHAVU_004G090900g	0.0	1723	915/971(94%)	940/971(96%)	1	1	2910	109..3018	
VuMYB64	3E-23	108	-3	52/103 (50%)	65/103 (62%)	1/103 (0%)	1298	ESW28753.1	Phaseolus vulgaris	hypothetical protein PHAVU_002G015100g	3,00E-168	498	283/302(94%)	292/302(96%)	-1	-1	936	1086..2021	
VuMYB65	5E-22	103	3	49/103 (47%)	64/103 (61%)	1/103 (0%)	988	ESW32880.1	Phaseolus vulgaris	hypothetical protein PHAVU_001G025200g	9,00E-146	438	241/275(88%)	247/275(89%)	-2	-2	822	1104..1925	
VuMYB66	1E-46	185	1	82/116 (70%)	98/116 (83%)		1287	ESW35375.1	Phaseolus vulgaris	hypothetical protein PHAVU_001G229900g	5,00E-177	519	285/330(86%)	300/330(90%)	3	3	987	690..1676	
VuMYB68	3E-23	108	2	52/103 (50%)	65/103 (62%)	1/103 (0%)	1563	ESW27881.1	Phaseolus vulgaris	hypothetical protein PHAVU_003G240200g	5,00E-177	514	279/291(96%)	281/291(96%)	-1	-1	951	648..1598	
VuMYB70	1E-32	140	2	81/198 (40%)	112/198 (55%)		1399	ESW14862.1	Phaseolus vulgaris	hypothetical protein PHAVU_007G023600g	1,00E-154	454	221/241(92%)	230/241(95%)	1	1	726	235..960	
VuMYB71	6E-45	180	2	84/116 (72%)	93/116 (79%)		1633	ESW11528.1	Phaseolus vulgaris	hypothetical protein PHAVU_008G038000g	0.0	566	351/386(91%)	358/386(92%)	-2	-2	948	435..1382	
VuMYB72	5E-59	226	-3	119/262 (45%)	154/262 (58%)	16/262 (6%)	1611	ESW15860.1	Phaseolus vulgaris	hypothetical protein PHAVU_007G108500g	1,00E-172	500	240/267(90%)	249/267(93%)	3	3	777	267..1043	
VuMYB73	4,00E-052	204	-1	88/118 (74%)	103/118 (86%)		1589	ESW19207.1	Phaseolus vulgaris	hypothetical protein PHAVU_006G105200g	1,00E-132	396	195/214(91%)	203/214(94%)	-1	-1	645	534..1178	
VuMYB76	4E-46	184	3	82/127 (64%)	97/127 (75%)		1293	ESW10181.1	Phaseolus vulgaris	hypothetical protein PHAVU_009G187700g	0.0	556	297/323(92%)	303/323(93%)	-1	-1	960	262..1221	
VuMYB77	8E-26	116	2	60/124 (48%)	77/124 (61%)	5/124 (4%)	752	ESW08649.1	Phaseolus vulgaris	hypothetical protein PHAVU_009G062700g	2,00E-121	365	174/191(91%)	181/191(94%)	-1	-1	1059	194..1252	
VuMYB78	1E-41	169	2	79/117 (67%)	92/117 (78%)		1075	ESW23221.1	Phaseolus vulgaris	hypothetical protein PHAVU_004G028500g	0.0	520	277/325(85%)	293/325(90%)	-2	-2	963	192..1154	
VuMYB79	3E-57	221	-3	113/241 (46%)	138/241 (56%)	13/241 (5%)	1276	XP_003535693.1	Glycine max	PREDICTED: myb-related protein Myb4	5,00E-119	360	201/264(76%)	216/264(81%)	2	2	759	209..967	
VuMYB80	3E-58	224	3	99/126 (78%)	115/126 (90%)		1107	ESW34538.1	Phaseolus vulgaris	hypothetical protein PHAVU_001G161000g	5,00E-149	437	263/295(89%)	275/295(93%)	3	3	885	174..1058	
VuMYB86	3E-48	191	-2	84/116 (72%)	97/116 (83%)		837	XP_003527903.1	Glycine max	PREDICTED: transcription factor MYB39-like	0.0	556	280/311(90%)	292/311(93%)	2	2	960	65..1024	
VuMYB88	4E-49	194	-1	86/116 (74%)	98/116 (84%)		1139	ESW04616.1	Phaseolus vulgaris	hypothetical protein PHAVU_011G110500g	1,00E-173	498	275/319(86%)	284/319(89%)	-1	-1	921	126..1046	
VuMYB89	3E-22	104	-2	50/104 (48%)	66/104 (63%)	1/104 (0%)	951	ESW08709.1	Phaseolus vulgaris	hypothetical protein PHAVU_009G068000g	3,00E-112	340	202/248(81%)	217/248(87%)	-3	-3	789	188..976	
VuMYB91	3E-33	141	1	61/115 (53%)	86/115 (74%)	1/115 (0%)	970	ESW11530.1	Phaseolus vulgaris	hypothetical protein PHAVU_008G038200g	3,00E-76	220	124/162(77%)	135/162(83%)	1	1	498	403..900	
VuMYB92	7E-32	136	2	61/101 (60%)	74/101 (72%)		894	ESW12445.1	Phaseolus vulgaris	hypothetical protein PHAVU_008G113300g	2,00E-94	293	165/182(91%)	172/182(94%)	1	1	477	481..957	
VuMYB94	4E-46	184	1	83/116 (71%)	95/116 (81%)		767	XP_003555012.1	Glycine max	PREDICTED: myb-related protein 315-like	5,00E-140	409	219/280(78%)	239/280(85%)	1	1	843	13..855	
VuMYB96	9E-33	139	3	61/115 (53%)	85/115 (73%)		918	ESW10665.1	Phaseolus vulgaris	hypothetical protein PHAVU_009G228200g	2,00E-78	248	140/177(79%)	151/177(85%)	3	3	183	708..890	
VuMYB101	8E-39	159	2	74/134 (55%)	92/134 (68%)		1355	ESW06819.1	Phaseolus vulgaris	hypothetical protein PHAVU_010G079300g	0.0	592	332/449(74%)	355/449(79%)	-2	-2	482	263..744	
VuMYB102	9E-20	96.7	2	47/106 (44%)	69/106 (64%)	1/106 (0%)	2639	ESW12310.1	Phaseolus vulgaris	hypothetical protein PHAVU_008G102000g	0.0	1078	538/560(96%)	545/560(97%)	1	1	1170	1926..3095	
VuMYB103	7E-43	173	-1	79/116 (68%)	94/116 (80%)		1992	ESW17051.1	Phaseolus vulgaris	hypothetical protein PHAVU_007G206200g	0.0	689	345/403(86%)	357/403(88%)	1	1	1683	151..1833	
VuMYB104	4E-18	91.7	3	56/149 (37%)	73/149 (48%)	1/149 (0%)	1617	ESW22601.1	Phaseolus vulgaris	hypothetical protein PHAVU_005G166300g	0.0	627	322/355(91%)	332/355(93%)	3	3	1164	187..1350	
VuMYB105	2E-49	195	3	88/139 (63%)	110/139 (78%)	6/139 (4%)	1065	ESW21307.1	Phaseolus vulgaris	hypothetical protein PHAVU_005G060000g	0.0	616	338/405(83%)	365/405(90%)	3	3	1065	522..1586	
VuMYB106	1E-36	152	3	68/105 (64%)	80/105 (75%)		1671	ESW07033.1	Phaseolus vulgaris	hypothetical protein PHAVU_010G096400g	1,00E-126	384	223/271(82%)	237/271(87%)	-3	-3	1224	207..1430	
VuMYB107	3E-46	184	-1	82/116 (70%)	97/116 (82%)		1667	ESW27576.1	Phaseolus vulgaris	hypothetical protein PHAVU_003G214300g	5,00E-117	360	227/244(93%)	231/244(94%)	-1	-1	849	794..1642	
VuMYB109	4E-39	160	-3	74/117 (63%)	90/117 (76%)	1/117 (0%)	1545	ESW08789.1	Phaseolus vulgaris	hypothetical protein PHAVU_009G075000g	0.0	555	308/347(89%)	320/347(92%)	-1	-1	1014	148..1161	
VuMYB110	0,00005	47.8	-3	28/59 (47%)	36/59 (60%)	1/59 (1%)	1500	NP_001240048.1	Glycine max	uncharacterized protein LOC100783949	2,00E-164	479	271/310(87%)	278/310(89%)	3	3	1029	272..1300	
VuMYB112	4E-19	94.4	-3	45/97 (46%)	59/97 (60%)	1/97 (1%)	300	XP_003527964.1	Glycine max	PREDICTED: transcription factor LAF1-like	7,00E-169	491	285/342(83%)	301/342(88%)	-2	-2	924	63..986	
VuMYB114	3E-33	141	1	63/128 (49%)	85/128 (66%)		1268	ESW35200.1	Phaseolus vulgaris	hypothetical protein PHAVU_001G215100g	4,00E-124	371	208/254(82%)	215/254(84%)	-1	-1	195	106..300	
VuMYB115	3E-19	95.5	-3	42/75 (56%)	55/75 (73%)		1131	ESW32232.1	Phaseolus vulgaris	hypothetical protein PHAVU_002G304400g	3,00E-87	275	165/267(62%)	172/267(64%)	-1	-1	831	261..1091	
VuMYB118	8E-45	179	-3	80/118 (67%)	93/118 (78%)		1081	ESW19324.1	Phaseolus vulgaris	hypothetical protein PHAVU_006G114800g	6,00E-135	397	228/281(81%)	249/281(88%)	-3	-3	513	461..973	

Anexo 3: Normas para submissão na revista *Genetics and Molecular Biology*.

SUBMISSION OF PAPERS

There is a publication charge for manuscripts once they are accepted. For price information, exemptions and waiver policies, please consult the journal homepage <http://www.gmb.org.br>.

1. Manuscripts must be submitted through our online submission platform hosted at: <http://mc.manuscriptcentral.com/gmb>

The cover letter should be addressed to:

Carlos C. F. Menck, Editor-in-Chief, Genetics and Molecular Biology

2. For submission the following instructions must be observed:

- The manuscript must be submitted by the Corresponding Author. This is the person who will also check the page proofs, and arranges for any payment that may incur during the editorial process.
- Entering the following metadata is required: (i) the manuscript title, (ii) a short running title (max. 35 characters), (iii) the Abstract, and (iv) up to five keywords. All these items must be exactly the same as those figuring in the first two pages of the manuscript file. Furthermore, a cover letter addressed to the Editor is required. It must be edited by the corresponding author inserted within the reserved data field.
- Statements are required informing that the data have not been published and are not under consideration elsewhere, and that all authors have approved the submission of the manuscript. Furthermore, possible conflicts of interest (e.g. due to funding, consultancies) must also be disclosed.
- The names of all co-authors, including institutional affiliations and e-mail addresses must be entered, as contact information for the Editorial Office.
- In the referee suggestions field, up to five reviewer names can be entered by the author(s); valid e-mail contact addresses for these are required, in case they are selected by the editor. These suggestions can be made separately as preferred and not-preferred reviewer(s).
- Files must be uploaded separately and identified according to file types. The main text file must include references and, if applicable, figure legends, which must be typed on a separate page following the References and Internet Resources sections. Each table, figure and element containing supplementary material must be saved and uploaded in a separate file. Formats for text and tables are Word or RTF in Windows platform. Figures should be in TIFF or JPEG formats (see detailed instructions in 3.1.h).
- Manuscripts including photos or any other identifiable data of human subjects must be accompanied by a copy of the signed consent by the individual or his/her guardian.

Failure to adhere to these guidelines can delay the handling of your contribution and manuscripts may be returned before being reviewed. Special attention should be given to the structuring of the manuscript and correct language usage. These are important factors in the smooth running of the editorial and peer-review process, and can result in faster publication.

3. Categories of Contribution

3.1. Research Articles

Manuscripts must be written in English in double-spaced, 12-point type throughout; marked with consecutive line and page numbers, beginning with the cover page. The following elements must start on a new page and be ordered as they are listed below:

a) **The title page** must contain: a concise and informative title; the authors' names (first name at full length); the authors' institutional affiliation, including department, institution, city, state or province, and country; different affiliations indicated with superscript Arabic numbers; a short running title of up to 35 characters (including spaces); up to five key words; the corresponding author's name, full postal, and email address.

b) **The Abstract** must be a single paragraph that does not exceed 200 words and summarizes the main results and conclusions of the study. It should not contain references.

c) **The text** must be as succinct as possible. *Text citations*: articles should be referred to by authors' surnames and date of publication; citations with two authors must include both names; in citations with three or more authors, name the first author and use *et al.* List two or more references in the same citation in chronological order, separated by semi-colons. When two or more works in a citation were published in the same year, list them alphabetically by the first author surname. For two or more works by the same author(s) in a citation, list them chronologically, with the years separated by commas. (Example: Freire-Maia *et al.*, 1966a, 1966b, 2000). Only articles that are published or in press should be cited. In the case of personal communications or unpublished results, all contributors must be listed by initials and last name (*et al.* should not be used). *Numbers*: In the text, numbers nine or less must be written out except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numerals for numbers larger than nine. *Binomial Names*: Latin names of genera, species and infraspecific taxa must be printed in italics; names of orders and families should appear in the Title and also when first mentioned in the text. URLs for programs, data or other sources should be listed in the Internet Resources Section, immediately following the References Section, not in the text. The text includes the following elements: *Introduction* - Description of the background that led to the study. *Material (or Subjects) and Methods* - Details relevant to the conduct of the study. Statistical methods should be explained at the end of this section. *Results* - Undue repetition in text and tables should be avoided. Statistical analyses should be presented as complete as possible, i.e. not only *P*-values should be shown, but also all other test variables required for full appreciation of the results by the reviewers and readers. Comments on relevance of results are appropriate but broader discussion should be part of the Discussion section. *Discussion* - The findings of the study should be placed in context of relevant published data. Ideas presented in other publications should not be discussed solely to make an exhaustive presentation. Some manuscripts may require different formats appropriate to their content.

d) **The Acknowledgments** must be a single paragraph that immediately follows the discussion and includes references to grant support.

e) **The References Section**: References must be ordered alphabetically by the first author surname; references with the same first author should be ordered as follows: first, as single author in chronological order; next, with only one more co-author in alphabetical order by the second author; and finally followed by references with more than two co-authors, in chronological order, independent of the second author surnames. In references with more than 10 authors only the first ten should be listed, followed by *et al.* Use standard abbreviations for journal titles as suggested by NCBI (<http://www.ncbi.nlm.nih.gov/journals/>). Only articles that are published or in press should be included in this section. Works submitted for publication but not yet accepted, personal communications and unpublished data must be cited within the text. "Personal communication" refers to information obtained from individuals other than the authors of the manuscript being submitted; "unpublished data" refers to data produced by at least one of the authors of the manuscript under consideration. Works of restricted circulation (e.g., theses not available in public databases, congress abstracts not published in regular journals or public databases) should not be listed in this section.

Sample journal article citation:

Breuer ME and Pavan C (1955) Behaviour of polytene chromosomes of *Rhynchosciara angelae* at different

stages of larval development. Chromosoma 7:371-386.
 Yonenaga-Yassuda Y, Rodrigues MT and Pellegrino KCM (2005) Chromosomal banding patterns in the eyelid-less microteiid lizard radiation: The X1X1X2X2:X1X2Y sex chromosome system in *Calyptommatus* and the karyotypes of *Psilophthalmus* and *Tretioscincus* (Squamata, Gymnophthalmidae). Genet Mol Biol 28:700-709.

Sample book citation:

Dobzhansky T (1951) Genetics and Origin of Species. 3rd edition. Columbia University Press, New York, 364 pp.

Sample chapter-in-book citation:

Crawford DC and Howard-Peebles PN (2005) Fragile X: From cytogenetics to molecular genetics. In: Gersen SL and Keagle MB (eds) The Principles of Clinical Cytogenetics. 2nd edition. Humana Press, New Jersey, pp 495-513.

Sample electronic article citation:

Gotzek D, Ross KG (2009) Current status of a model System: The gene *Gp-9* and its association with social organization in fire ants. PLoS One 4:e7713.

f) **Internet Resources Section:** this section should contain a list of URLs referring to data presented in the text, as well as software programs and other Internet resources used during data processing. Date of consultation must be stated.

Sample	Internet	resource	citation:
Online Mendelian Inheritance in Man (OMIM),	http://www.ncbi.nlm.nih.gov/OMIM		(September 4, 2009)
LEM Software,	http://dir.niehs.nih.gov/dirbb/weinbergfiles/hybrid_design.htm		(September 4, 2009)

g) **Tables:** must be in Word format prepared with the table tool (do not use space bar or tabulator). A concise title should be provided above the table. Tables must be numbered consecutively in Arabic numerals. Each column must have a title in the box head. Footnotes typed directly below the table should be indicated in lowercase superscript letters. Tables that are to appear in the printed version must be saved in Word format and not as figures, so that they can later be fitted during typesetting. Each table must be saved and uploaded as a separate file.

h) **Figures** must be numbered consecutively using Arabic numerals. Images should be in TIFF or JPEG format. Figures in Word, PowerPoint or Excel format cannot be published. Only sequence data can be presented in Word format. Journal quality reproduction will require grayscale resolution yielding 300 dpi, color figures should be at 600 dpi. These resolutions refer to the output size of the file, that is the size in which it will appear printed in the journal; if it is anticipated that images will be enlarged or reduced, the resolutions should be adjusted accordingly. Figures composed of several elements should be sent as a single panel, obeying the print size definitions of the journal (single or two columns width). Scanned figures should not be submitted. Color illustrations are accepted. Each figure/panel must be saved and uploaded as a separate file. When uploading, identify each illustration by the first author name and the number of the respective figure. Figure legends must be included at the end of the main text file and should be typed on a new page.

i) **Nomenclature:** Taxonomic names should be in accordance with current international standards. For rules concerning gene names and gene symbols, please see separate Instruction form.

j) **Sequences** may appear in text or in figure. DNA, RNA and protein sequences equal to or greater than 50 units must be entered into public databases and accession numbers must be provided upon acceptance of the article. Failure to do so will inadvertently delay publication.

k) **Data access:** reference should be made to availability of detailed data and materials used for reported

studies.

l) **Ethical issues:** Reports of experiments on live vertebrates must include a statement in the text that the institutional review board approved the work and the protocol number must be provided. For experiments involving human subjects, authors must also include a statement that informed consent was obtained from all subjects. If photos or any other identifiable data are included, a copy of the signed consent must be uploaded during manuscript submission.

m) **Supplementary Material:** Data that the authors consider of importance for completeness of a study, but which are too extensive to be included in the published version, can be submitted as Supplementary Material. At publication, this material will be made available together with the electronic version. In case a manuscript contains such material, it should be appropriately identified within the text file. Supplementary material in tables should be identified as Table S1, Table S2, etc., in case of figures they should be named accordingly, Figure S1, Figure S2. In addition, a list of this material should be presented at the end of the manuscript text file, containing the following statement:

Supplementary material - the following online material is available for this article:

- *Table S1 – < short title >*
- *Figure S1 – < short title >*

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

3.2 Short Communications

Short Communications present brief observations that do not warrant full-length articles. They should not be considered preliminary communications;

- should be 15 or fewer typed pages in double spaced 12-point type, including literature cited;
- should include an Abstract;
- but no further subdivision, with introduction, material and methods, results and discussion; all in a single section and without headers.
- up to four items (tables and/or figures) may be submitted;

Note: The title page, abstract and reference section format is that of a full-length Research Article. For Supplementary Material see instructions in item 3.1.m.

3.3 Letters to the Editor

Relate or respond to recent published items in the journal. Discussions of political, social and ethical issues of interest to geneticists are also welcome in this form.

3.4 Review Articles

Review Articles are welcome. The Editor must be contacted prior to submission. Please, provide an Abstract and a list of your recent publications in the area.

3.5 Book Reviews

Publishers are invited to submit books on Genetics, Evolution and related disciplines, for review in the journal. Aspiring reviewers may propose writing a review.

3.6 History, Story and Memories

These are accounts on historical aspects of Genetics relating to Brazil.

4. Articles accepted for publication

Once an article is accepted, the Editorial Office will send it to copy editor for language and technical corrections. If major corrections were proposed, the manuscript with the highlighted corrections will be returned to the corresponding author for approval. The final version approved by the authors must be free of any text/correction markings when returned to the Editorial Office.

After typesetting, page proofs will be sent to the corresponding author. Changes made to page proofs, apart from typesetting errors, will be charged to the authors. Notes added in proof require Editorial approval. Together with the proofs, a form of consent to publish and transfer of copyright is sent to the corresponding other. The latter will have sign this form, also on behalf of any co-authors, and send it by fax to the Editorial Office.

5. Reprints

Reprints are free of charge and will be provided as a pdf-file.

Anexo 4: Normas para submissão na revista *International Journal of Molecular Sciences*.

Manuscript Submission Overview

Types of Publications

IJMS has no restrictions on the length of manuscripts, provided that the text is concise and comprehensive. Full experimental details must be provided so that the results can be reproduced by other groups. *IJMS* encourages authors to publish all experimental controls and full datasets as supplementary files (please read the guidelines about [Supplementary Materials](#) carefully and [references to unpublished data](#)).

The different types of articles published in *IJMS* are indicated in the first section of the [Aims & Scope](#). The main types are:

- *Articles*: research manuscripts report new evidence or new conclusions which have neither been published before nor are under consideration for publication in another journal. MDPI considers all original research manuscripts provided that the work reports scientifically sound experiments and provides a substantial amount of new information. We strongly recommend authors not to unnecessarily divide their work into several related manuscripts.
- *Short communications* of preliminary, but significant, results will also be considered.
- *Reviews*: review manuscripts provide concise and precise updates on the latest progress made in a given area of research.
- *Conference Papers*: Expanded and high quality conference papers are also considered in *IJMS* if they fulfill the following requirements: (1) the paper should be expanded to the size of a research article; (2) the conference paper should be cited and noted on the first page of the paper; (3) if the authors do not hold the copyright to the published conference paper, authors should seek the appropriate permission from the copyright holder; (4) authors are asked to disclose that it is conference paper in their cover letter and include a statement on what has been changed compared to the original conference paper.

Submission Process

Manuscripts for *IJMS* should be submitted online at susy.mdpi.com. The submitting author, who is generally the corresponding author, is responsible for the manuscript during the submission and peer-review process. The submitting authors must ensure that all co-authors have been included in the author list (read the [criteria to qualify for authorship](#)) and that they all have read and approved

the submitted version of the manuscript. To submit your manuscript, [register and log in](#) to this website. Once you are registered, [click here to go to the submission form for IJMS](#). All co-authors can see the manuscript details in the submission system, if they register and log in using the e-mail address provided during manuscript submission.

Accepted File Formats

Authors must use the [Microsoft Word template](#) or [LaTeX template](#) to prepare their manuscript. Using the template file will substantially shorten the time to complete copy-editing and publication of accepted manuscripts. Accepted file formats are:

- **Microsoft Word:** Manuscripts prepared in Microsoft Word must be converted into a single file before submission. When preparing manuscripts in Microsoft Word, the [IJMS Microsoft Word template file](#) must be used. Please insert your graphics (schemes, figures, etc.) in the main text after the paragraph of its first citation.
- **LaTeX:** Manuscripts prepared in LaTeX must be collated into one ZIP folder (include all source files and images, so that the Editorial Office can recompile the submitted PDF). When preparing manuscripts in LaTeX, please use the [IJMS LaTeX template files](#). You can now also use the online application [writeLaTeX](#) to submit articles directly to IJMS. The MDPI LaTeX template file should be selected from the [writeLaTeX template gallery](#).

Cover Letter

A cover letter must be included with each manuscript submission. It should be concise and explain why the content of your paper is significant, placing your findings in the context of existing work and why it fits the scope of the journal. Please confirm that neither the manuscript nor any parts of its content are currently under consideration or published in another journal. Any prior submissions of the manuscript to MDPI journals must be acknowledged. The names of proposed and excluded reviewers should be provided in the submission system, not in the cover letter.

Note for Authors Funded by the National Institutes of Health (NIH)

This journal automatically deposits papers to PubMed Central after publication of an issue. Authors do not need to separately submit their papers through the NIH Manuscript Submission System (NIHMS, <http://nihms.nih.gov/>).

Preparation of a Manuscript

General Considerations

- **Research manuscripts** should comprise:
 - [Front matter](#): Title, Author list, Affiliations, Abstract, Keywords
 - [Research manuscript sections](#): Introduction, Results, Discussion, Conclusions (optional), Materials and Methods, Supplementary Materials
 - [Back matter](#): Acknowledgments, Author Contributions, Conflict of Interests, References.
- **Review manuscripts** should comprise the [front matter](#), literature review sections and the [back matter](#). The template file can also be used to prepare the front and back matter of your review manuscript. It is not necessary to follow the remaining structure.
- **Abstract Graphic:** Authors are encouraged to provide a graphical abstract to display on the website alongside the textual abstract. It should be a self-explanatory snapshot of your article giving a view on its rationale, study design, and/or conclusions. The graphic should not exceed 550 pixels. When prepared in Adobe Photoshop or Microsoft PowerPoint, the frame should be 5–15 cm in width and height. The text should be kept to a minimum and

the font size comprised between 10 pt and 14 pt to ensure readability. The graphic should be provided as a JPG, TIFF, PNG or GIF file.

- **"Data not shown"** should be avoided in research manuscripts. We encourage our authors to publish all observations related to the submitted manuscript as Supplementary Materials. **"Unpublished data"** intended for publication in a different manuscript, *i.e.*, in a manuscript that is either planned, **"in preparation"** or that have been **"submitted"** but not yet accepted, should be cited in the text and a reference should be added in the References section. **"Personal Communications"** should also be cited in the text and reference added in the References section. (see also the MDPI reference list and citations style guide).
- **Abbreviations** should be defined in parentheses the first time they appear in the abstract, main text and in figure captions.
- **SI Units** (International System of Units) should be used for this journal. Imperial, US customary and other units should be converted to SI units whenever possible before submission of a manuscript to the journal.
- **Accession numbers** of RNA, DNA and protein sequences used in the manuscript should be provided in the Materials and Methods section. Please also read the Guidelines for Deposition of Sequences and of Expression Data
- **Equations:** If you are using Word, please use either the Microsoft Equation Editor or the MathType add-on in your paper. Equations should be editable by the editorial office and not appear in a picture format.
- **Supplementary Materials and Research Data:** To maintain the transparency and reproducibility of research results, authors are encouraged to make their experimental and research data openly available either by depositing into data repositories or by publishing the data and files as "Supplementary Materials". Large datasets and files should be deposited in specialized data repositories. Small datasets, spreadsheets, images, video sequences, conference slides, software source code, *etc.* can be uploaded as "Supplementary Files" during the manuscript submission process. The supplementary files will also be made available to the referees during the peer-review process and be published online alongside the manuscript. Please read the information about Supplementary Materials and Data Deposit for additional guidelines.

Front Matter

These sections should appear in all manuscript types

- **Title:** The title of your manuscript should be concise, specific and relevant. When gene or protein names are included, the abbreviated name rather than full name should be used.
- **Authors List and Affiliations:** Authors' full first and last names must be provided. The initials of any middle names can be added. The PubMed/MEDLINE standard format is used for affiliations: complete address information including city, zip code, state/province, country, and all email addresses. At least one author should be designated as corresponding author, and his or her email address and other details should be included at the end of the affiliation section. Please read the criteria to qualify for authorship.
- **Abstract:** The abstract should be a total of about 200 words maximum. The abstract should be a single paragraph and should follow the style of structured abstracts, but without headings: 1) Background: Place the question addressed in a broad context and highlight the purpose of the study; 2) Methods: Describe briefly the main methods or treatments applied; 3) Results: Summarize the article's main findings; and 4) Conclusion: Indicate the main conclusions or interpretations. The abstract should be an objective representation of the article: it must not contain results which are not presented and substantiated in the main text and should not exaggerate the main conclusions.
- **Keywords:** Three to ten pertinent keywords need to be added after the abstract. We recommend that the keywords are specific to the article, yet reasonably common within the subject discipline.

Research Manuscript Sections

- **Introduction:** The introduction should briefly place the study in a broad context and highlight why it is important. It should define the purpose of the work and its significance. The current state of the research field should be reviewed carefully and key publications should be cited. Please highlight controversial and diverging hypotheses when necessary. Finally, briefly mention the main aim of the work and highlight the main conclusions. As far as possible, please keep the introduction comprehensible to scientists outside your particular field of research.
- **Results:** This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.
- **Discussion:** This section may be divided by subheadings. Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.
- **Conclusions:** This section is not mandatory, but can be added to the manuscript if the discussion is unusually long or complex.
- **Materials and Methods:** This section should be divided by subheadings. Materials and Methods should be described with sufficient details to allow others to replicate and build on published results. Please note that publication of your manuscript implicates that you must make all materials, data, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited.
 - Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication.
 - Research manuscripts using human or animal subjects, tissues, field samples or cell lines must include research ethics statements. Please read the [Ethical Research Guidelines](#).
- **Supplementary Materials:** This section should be included when supplementary information is published online alongside the manuscript. Please indicate the name and title of each supplementary file as follows **Figure S1:** title, **Table S1:** title, etc.

Back Matter

- **Acknowledgments:** All sources of funding of the study should be disclosed. Please clearly indicate grants that you have received in support of your research work. Clearly state if you received funds for covering the costs to publish in open access. Note that some funders will not refund article processing charges (APC) if the funder and grant number are not clearly identified in the paper. Funding information can be entered separately into the submission system by the authors during submission of their manuscript. Such funding information, if available, will be deposited to [FundRef](#) if the manuscript is finally published.
- **Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "X.X. and Y.Y. conceived and designed the experiments; X.X. performed the experiments; X.X. and Y.Y. analyzed the data; W.W. contributed reagents/materials/analysis tools; Y.Y. wrote the paper. **Authorship must be limited to those who have contributed substantially to the work reported. Please read the section concerning the [criteria to qualify for authorship](#) carefully.**
- **Conflicts of Interest:** Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. If there is no conflict of interest, please state "The authors declare no conflict of interest." Any role of the funding sponsors in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The founding sponsors had no role in the design of the study;

in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results”.

- **References:** References must be numbered in order of appearance in the text (including tables and legends) and listed individually at the end of the manuscript. We recommend preparing the references with a bibliography software package, such as EndNote, ReferenceManager or Zotero to avoid typing mistakes and duplicated references.

Citations and References in Supplementary files are permitted provided that they also appear in the main text and in the reference list.

- In the text, reference numbers should be placed in square brackets [], and placed before the punctuation; for example [1], [1–3] or [1,3]. For embedded citations in the text with pagination, use both parentheses and brackets to indicate the reference number and page numbers; for example [5] (p. 10). or [6] (pp. 101–105).
- The Reference list should include the full title as recommended by the ACS style guide. The style file for endnote, MDPI.ens, can be found at <http://endnote.com/downloads/style/mdpi>
- References should be described as follows depending on the type of work:
 - **Journal Articles:**
1. Author 1, A.B.; Author 2, C.D. Title of the article. *Abbreviated Journal Name* **Year**, *Volume*, page range, DOI or other identifier. Available online: URL (accessed on Day Month Year).
 - **Books and Book Chapters:**
2. Author 1, A.; Author 2, B. *Book Title*, 3rd ed.; Publisher: Publisher Location, Country, Year; pp. 154–196.
3. Author 1, A.; Author 2, B. Title of the chapter. In *Book Title*, 2nd ed.; Editor 1, A.; Editor 2, B., Eds.; Publisher: Publisher Location, Country, Year; Volume 3, pp. 154–196.
 - **Unpublished work, submitted work, personal communication:**
4. Author 1, A.B.; Author 2, C. Title of Unpublished Work. status (unpublished; manuscript in preparation).
5. Author 1, A.B.; Author 2, C. Title of Unpublished Work. *Abbreviated Journal Name* stage of publication (under review; accepted; in press).
6. Author 1, A.B. (University, City, State, Country); Author 2, C. (Institute, City, State, Country). Personal communication, Year.
 - **Conference Proceedings:**
7. Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In *Title of the Collected Work* (if available), Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference; Editor 1, Editor 2, Eds. (if available); Publisher: City, Country, Year (if available); Abstract Number (optional), Pagination (optional).
 - **Thesis:**
8. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.
 - **Websites:**
9. Title of Site. Available online: URL (accessed on Day Month Year). Unlike published works, websites may change over time or disappear, so we encourage you create an archive of the cited website using a service such as WebCite. Archived websites should be cited using the link provided as follows:
10. Title of Site. URL (archived on Day Month Year).

See the Reference List and Citations Guide for more detailed information.

Preparing Figures, Schemes and Tables

- All figure files should be separately uploaded during submission.

- Figures and schemes must be provided at a sufficiently high resolution (minimum 1000 pixels width/height, or a resolution of 300 dpi or higher). All Figure file formats are accepted. However, TIFF, JPEG, EPS and PDF files are preferred.
- *IJMS* can publish multimedia files in articles or as supplementary materials. Please get in touch with the Editorial office for further information.
- All Figures, Schemes and Tables should also be inserted into the main text close to their first citation and must be numbered following their number of appearance (Figure 1, Scheme I, Figure 2, Scheme II, Table 1, etc.).
- All Figures, Schemes and Tables should have a short explanatory title and a caption placed above it.
- All table columns should have an explanatory heading. To facilitate the copy-editing of larger tables, smaller fonts may be used, but in no less than 8 pt. in size. Authors should use the Table option of Microsoft Word to create tables.
- For multi-panel figures, the file must contain all data in one file. For tips on creating multi-panel figures, please read the helpful advice provided by [L2 Molecule](#).
- Authors are encouraged to prepare figures and schemes in color (RGB at 8-bit per channel). Full color graphics will be published free of charge.

Qualification for Authorship

Authorship must include and be strictly limited to researchers who have substantially contributed to the reported work. To qualify for authorship, a researcher should have made a substantial contribution to the design of the study, or to the production, analysis or interpretation of the results. Authors should also have been involved in the preparation and have approved the submitted manuscript. Those who contributed to the work but do not qualify for authorship should be listed in the acknowledgments. According to the Committee on Publication Ethics ([COPE](#)) standard, to which this journal adheres, "all authors should agree to be listed and should approve the submitted and accepted versions of the publication. Any change to the author list should be approved by all authors including any who have been removed from the list. The corresponding author should act as a point of contact between the editor and the other authors and should keep co-authors informed and involve them in major decisions about the publication (e.g. answering reviewers' comments)." [1]

1. Wager, E.; Kleinert, S. Responsible research publication: international standards for authors. A position statement developed at the 2nd World Conference on Research Integrity, Singapore, July 22-24, 2010. In *Promoting Research Integrity in a Global Environment*; Mayer, T., Steneck, N., eds.; Imperial College Press / World Scientific Publishing: Singapore; Chapter 50, pp. 309-16.

Potential Conflicts of Interest

It is the authors' responsibility to identify and declare any personal circumstances or interests that may be perceived as inappropriately influencing the representation or interpretation of clinical research. If there is no conflict, please state "The authors declare no conflict of interest". This should be conveyed in a separate "Conflicts of Interest" section preceding the "References" sections at the end of the manuscript. Financial support for the study must be fully disclosed in the "Acknowledgments" section.

Editorial Procedure and Peer-Review

Initial Checks

All submitted manuscripts received by the Editorial Office will be checked by a professional in-house *Managing Editor* to determine whether it is properly prepared and whether the manuscript follows the ethical policies of the journal, including those for human and animal experimentation.

Manuscripts that do not fit the journal or are not in line with our ethical policy may be rejected before peer-review. Manuscripts that are not properly prepared will be returned to the authors for revision and resubmission. The *Managing Editor* will consult the journals' *Editor-in-Chief*, the *Guest Editor* or an *Editorial Board member* to determine whether the manuscript fits the scope of the journal and whether it is scientifically sound. No judgment on the significance or potential impact of the work will be made at this stage.

Peer-Review

Once a manuscript passes the initial checks, it will be assigned to at least two independent experts for peer-review. A single-blind peer-review process is applied, where authors' names are revealed to reviewers. In-house assistant editors generally invite experts recommended by the *Editor-in-Chief* or identified by literature searches. These experts may also include Editorial Board members and Guest Editors of the journal. Potential reviewers suggested by the authors may also be considered. Reviewers should not have published with any of the co-authors during the past five years and should not currently work or collaborate with one of the institutes of the co-authors of the submitted manuscript.

Editorial Decision and Revision

Based on the comments and advices of the peer-reviewers, an external editor – usually the *Editor-in-Chief* or a *Guest Editor* – will make a decision to accept, reject, or to ask authors to revise the manuscript.

For *Minor Revisions* the authors will have one week to resubmit their revised manuscript. For *Major Revisions* the authors will have two weeks to resubmit their revised manuscript. However, authors should contact the editorial office if extended revision time is anticipated.

Author Appeals

Authors may appeal a rejection by sending an e-mail to the Editorial Office of the journal. The appeal must provide a detailed justification, including point-by-point responses to the reviewers' and/or Editor's comments. The *Managing Editor* of the journal will forward the manuscript and relating information (including the identities of the referees) to an Editorial Board member who was not involved in the initial decision-making process. If no appropriate Editorial Board member is available, the editor will identify a suitable external scientist. The Editorial Board member will be asked to give an advisory recommendation on the manuscript and may recommend acceptance, further peer-review, or uphold the original rejection decision. A reject decision at this stage will be final and cannot be revoked.

Production and Publication

Once accepted, the manuscript will undergo professional copy-editing, English editing, proofreading by the authors, final corrections, pagination, and, publication on the www.mdpi.com website.

Suggestion of Reviewers

During the submission process, authors have the possibility to suggest potential reviewers with the appropriate expertise to review the manuscript. The editors will not necessarily approach these referees. Please provide detailed contact information (address, homepage, phone, e-mail address). The proposed referees should neither be current collaborators of the co-authors nor have published with any of the co-authors of the manuscript within the last five years. Proposed reviewers should be from different institutions to the authors. You may identify appropriate Editorial

Board members of the journal as potential reviewers. You may also suggest reviewers from among the authors that you frequently cite in your paper.

English Corrections

This journal is published in English. To facilitate proper peer-reviewing of your manuscript, it is essential that it is submitted in grammatically correct English. If you are not a native English speaker, we strongly recommend that you have your manuscript professionally edited before submission or read by a native English-speaking colleague. Professional editing will mean that reviewers and future readers are better able to read and assess the content of your manuscript. For additional information see the [English Editing Guidelines for Authors](#).

Publication Ethics Statement

IJMS is a member of the Committee on Publication Ethics ([COPE](#)). We fully adhere to its [Code of Conduct](#) and to its [Best Practice Guidelines](#).

The editors of this journal take the responsibility to enforce a rigorous peer-review process together with strict ethical policies and standards to ensure to add high quality scientific works to the field of scholarly publication. Unfortunately, cases of plagiarism, data falsification, image manipulation, inappropriate authorship credit, and the like, do arise. The editors of *IJMS* take such publishing ethics issues very seriously and are trained to proceed in such cases with a zero tolerance policy.

Authors wishing to publish their papers in *IJMS* are asked to abide to the following rules:

- Any facts that might be perceived as a possible conflict of interest of the author(s) must be disclosed in the paper prior to submission.
- Authors should accurately present their research findings and include an objective discussion of the significance of their findings.
- Data and methods used in the research need to be presented in sufficient detail in the paper, so that other researchers can replicate the work.
- Raw data should preferably be publicly deposited by the authors before submission of their manuscript. Authors need to at least have the raw data readily available for presentation to the referees and the editors of the journal, if requested. Authors need to ensure appropriate measures are taken so that raw data is retained in full for a reasonable time after publication.
- Simultaneous submission of manuscripts to more than one journal is not tolerated.
- Republishing content that is not novel is not tolerated (for example, an English translation of a paper that is already published in another language will not be accepted).
- If errors and inaccuracies are found by the authors after publication of their paper, they need to be promptly communicated to the editors of this journal so that appropriate actions can be taken. Please refer to our [policy regarding publication of publishing addenda and corrections](#).
- Your manuscript should not contain any information that has already been published. If you include already published figures or images, please obtain the necessary permission from the copyright holder to publish under the CC-BY license.
- Plagiarism, data fabrication and image manipulation are not tolerated.
 - **Plagiarism is not acceptable** in *IJMS* submissions.

Plagiarism includes copying text, ideas, images, or data from another source, even from your own publications, without giving any credit to the original source.

Reuse of text that is copied from another source must be between quotes and the original source must be cited. If a study's design or the manuscript's structure or language has been inspired by previous works, these works must be explicitly cited.

If plagiarism is detected during the peer review process, the manuscript may be rejected. If plagiarism is detected after publication, we may publish a correction or retract the paper.

- **Image files must not be manipulated or adjusted in any way** that could lead to misinterpretation of the information provided by the original image.

Irregular manipulation includes: 1) introduction, enhancement, moving, or removing features from the original image; 2) grouping of images that should obviously be presented separately (e.g., from different parts of the same gel, or from different gels); or 3) modifying the contrast, brightness or color balance to obscure, eliminate or enhance some information.

If irregular image manipulation is identified and confirmed during the peer review process, we may reject the manuscript. If irregular image manipulation is identified and confirmed after publication, we may correct or retract the paper.

Our in-house editors will investigate any allegations of publication misconduct and may contact the authors' institutions or funders if necessary. If evidence of misconduct is found, appropriate action will be taken to correct or retract the publication. Authors are expected to comply with the best ethical publication practices when publishing with MDPI.

Supplementary Materials and Data Deposit

In order to maintain the integrity, transparency and reproducibility of research records, and to retain important chemical and structural information, authors are strongly encouraged to make their experimental and research data openly available either by depositing into data repositories or by publishing the data and files as supplementary information in this journal. Additional data and files can be uploaded as "Supplementary Files" during the manuscript submission process. The supplementary files will also be available to the referees as part of the peer-review process, although referees are not specifically asked to review these files. Accepted file formats include (but are not limited to):

- spectral data (NMR, IR, Raman, ESR, etc.) in JCAMP (JDX) format
- 3D coordinate structures (in PDB, MOL, XYZ or other common format)
- crystallographic information (in CIF format)
- data tables and spreadsheets (text files, MS Excel, OpenOffice, CSV, XML, etc.)
- text documents (text files, PDF, MS Word, OpenOffice, etc.; text documents will usually be converted to PDF files for publication)
- images (JPEG, PNG, GIF, TIFF, BMP, etc.)
- videos (AVI, MPG, QuickTime, etc.)
- executables (EXE, Java, etc.)
- software source code

Citations and References in Supplementary files are permitted provided that they also appear in the main text and in the reference list.

Large data sets and files should be deposited to specialized service providers (such as [Figshare](#)) or institutional/subject repositories, preferably those that use the [DataCite](#) mechanism. For a list of specialized repositories for the deposit of scientific and experimental data, please consult [databib.org](#) or [re3data.org](#). The data repository name, link to the data set (URL) and accession number, doi or handle number of the data set must be provided in the paper. The journal *Data* (ISSN 2306-5729) also accepts submissions of data set papers, and the publication of small data sets along with the paper, and/or software source codes is encouraged.

Guidelines for Deposition of Sequences and of Expression Data

New sequence information must be deposited to the appropriate database prior to submission of the manuscript. Accession numbers provided by the database should be included in the submitted manuscript. Manuscripts will not be published until the accession number is provided.

- **New nucleic acid sequences** must be deposited in one of the following databases: GenBank, EMBL, or DDBJ. Sequences should be submitted to only one database.
- **New high throughput sequencing (HTS) datasets**(RNA-seq, ChIP-Seq, degradome analysis, ...) must be deposited either in the GEO database or in the NCBI's Sequence Read Archive.
- **New microarray data** must be deposited either in GEO or ArrayExpress databases. The "Minimal Information About a Microarray Experiment" (MIAME) guidelines published by the Microarray Gene Expression Data Society must be followed.
- **New protein sequences** obtained by protein sequencing must be submitted to UniProt (submission tool SPIN).

<p<**All sequence names and the accession numbers provided by the databases should be provided in the Materials and Methods section of the article.**</p>

10. Curriculum vitae (Lattes)

Mitalle Karen da Silva Matos

Curriculum Vitae

Formação acadêmica/titulação

2013 Mestrado em Genética.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Identificação de Fatores de Transcrição Relacionados com a Defesa em Leguminosas Cultivadas
Orientador: Ana Maria Benko-Iseppon
Bolsista do(a): Conselho Nacional de Desenvolvimento Científico e Tecnológico

Formação complementar

2013 - 2013 Curso de curta duração em PCR em Tempo Real: Princípios Básicos e Aplicações.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil

Prêmios e títulos

2014 Menção Honrosa, Sociedade Brasileira de Genética
No âmbito do Premio Jovem Geneticista do Nordeste.

Produção

Produção bibliográfica

Trabalhos publicados em anais de eventos (resumo)

1. ARAUJO, F. T., **MATOS, M. K. S.**, OLIVEIRA, A. R. S. *et al.*
CARACTERIZAÇÃO E ANCORAGEM DE GENE DE RESISTÊNCIA DO TIPO NBS-LRR EM SOJA [*Glycine max* (L.) Merr., Fabaceae] In: 20º Encontro de Genética do Nordeste, 2014, Campina Grande.

20º Encontro de Genética do Nordeste. , 2014.

2. ARAUJO, F. T., **MATOS, M. K. S.**, BEZERRA-NETO, J. P. *et al.*
DESENVOLVIMENTO DE MARCADORES DO TIPO RGA (RESISTANCE GENE ANALOG) EM FEIJÃO-CAUPI In: 20º Encontro de Genética do Nordeste, 2014, Campina Grande.

20º Encontro de Genética do Nordeste. , 2014.

3. **MATOS, M. K. S.**, SILVA, M. D., BEZERRA-NETO, J. P. *et al.*
DIVERSIDADE ESTRUTURAL 'in silico' DE CANDIDATOS A FATORES DE TRANSCRIÇÃO DA FAMÍLIA MYB NO TRANSCRIPTOMA DO FEIJÃO-CAUPI In: 20º Encontro de Genética do Nordeste, 2014, Campina Grande.

20º Encontro de Genética do Nordeste. , 2014.

4. OLIVEIRA, R. C. G., OLIVEIRA, W. D., ARAUJO, F. T. *et al.*
MARCADORES SSR POLIMÓRFICOS ENTRE GENÓTIPOS DE FEIJÃO-CAUPI CONTRASTANTES PARA TOLERÂNCIA À SECA In: XI Congresso Latino Americano de Botânica, LXV Congresso Nacional de Botânica e XXXIV Encontro Regional de Botânicos MG, BA, ES., 2014, Salvador.

XI Congresso Latino Americano de Botânica, LXV Congresso Nacional de Botânica e

XXXIV Encontro Regional de Botânicos MG, BA, ES. , 2014.

5. **MATOS, M. K. S., BEZERRA-NETO, J. P., AMORIM, L. L. B. et al.**
PROSPECÇÃO DE GENES MYB NO TRANSCRIPTOMA DO FEIJÃO-CAUPI E ANCORAGEM EM PSEUDOCROMOSSOMOS DE *Phaseolus vulgaris* L. In: 20º Encontro de Genética do Nordeste, 2014, Campina Grande.

20º Encontro de Genética do Nordeste. , 2014.

6. **MATOS, M. K. S., BEZERRA-NETO, J. P., Amorim, L.L.B et al.**
STRUCTURAL DOMAIN ANALYSIS AND 'IN SILICO' EXPRESSION OF R2R3-MYB SUBFAMILY IN COWPEA (*Vigna unguiculata* (L.) WALP.) TRANSCRIPTOME In: 60º Congresso Brasileiro de Genética, 2014, Guarujá.

ANAIS DO 60º CONGRESSO BRASILEIRO DE GENÉTICA. , 2014.

Apresentação de trabalho e palestra

1. **MATOS, M. K. S.**

Identificação de Fatores de Transcrição Relacionados com a Defesa Vegetal em Feijão-Caupi [*Vigna unguiculata* (L.) Walp.] e em Outras Plantas Cultivadas, 2014. (Conferência ou palestra, Apresentação de Trabalho)

2. **MATOS, M. K. S.**

Large Scale In Silico Identification of MYB Family Genes from Wheat Expressed Sequence Tags, 2014. (Conferência ou palestra, Apresentação de Trabalho)

3. **MATOS, M. K. S., BEZERRA-NETO, J. P., AMORIM, L. L. B. et al.**

Prospecção de Genes MYB no Transcriptoma do Feijão-Caupi e Ancoragem em Pseudocromossomos de *Phaseolus vulgaris* L., 2014. (Congresso, Apresentação de Trabalho)

4. **MATOS, M. K. S.**

The R2R3-MYB Transcription factors Gene Family in Maize, 2014. (Conferência ou palestra, Apresentação de Trabalho)

5. **MATOS, M. K. S.**

Global Expression Profiling of Transcription Factor Genes Provides New Insights Into Pathogenicity and Stress Responses in the Rice Blast Fungus, 2013. (Conferência ou palestra, Apresentação de Trabalho)

6. **MATOS, M. K. S.**

Plant bZIP Transcription Factors Responsive to Pathogens: A Review, 2013. (Conferência ou palestra, Apresentação de Trabalho)

Eventos

Eventos

Participação em eventos

1. **20º Encontro de Genética do Nordeste**, 2014. (Congresso)

Prospecção de Genes MYB no Transcriptoma do Feijão-Caupi e Ancoragem em Pseudocromossomos de *Phaseolus vulgaris* L.

2. Apresentação de Pôster / Painel no(a) **60º Congresso Brasileiro de Genética**, 2014. (Congresso) Structural domain analysis and 'in silico' expression of R2R3-MYB subfamily in cowpea (*Vigna unguiculata* (L.) Walp.) transcriptome.