

EXPRESSÃO DE microRNAs NA GERMINAÇÃO DE *Glycine max*
(L.) MERRILL SOB ESTRESSES ABIÓTICOS E
CARACTERIZAÇÃO DE UMA FAMÍLIA GÊNICA ENVOLVIDA NA
METILAÇÃO DO DNA DIRIGIDA POR RNA EM PLANTAS

PAULA MACHADO DE ARAÚJO RIBEIRO

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES – RJ
FEVEREIRO – 2024

EXPRESSÃO DE microRNAs NA GERMINAÇÃO DE *Glycine max*
(L.) MERRILL SOB ESTRESSES ABIÓTICOS E
CARACTERIZAÇÃO DE UMA FAMÍLIA GÊNICA ENVOLVIDA NA
METILAÇÃO DO DNA DIRIGIDA POR RNA EM PLANTAS

PAULA MACHADO DE ARAÚJO RIBEIRO

Tese apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para a obtenção do título de Doutora em Biociências e Biotecnologia.

Orientadora: Prof.^a Dr.^a Clícia Grativol Gaspar de Matos

Campos dos Goytacazes – RJ
Fevereiro – 2024

FICHA CATALOGRÁFICA

UENF - Bibliotecas

Elaborada com os dados fornecidos pela autora.

R484

Ribeiro, Paula Machado de Araújo.

"Expressão de microRNAs na germinação de *Glycine max* (L.) Merrill sob estresses abióticos e caracterização de uma família gênica envolvida na metilação do DNA dirigida por RNA em plantas". / Paula Machado de Araújo Ribeiro. - Campos dos Goytacazes, RJ, 2024.

119 f. : il.

Inclui bibliografia.

Tese (Doutorado em Biociências e Biotecnologia) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, 2024.

Orientadora: Clécia Grativol Gaspar.

1. pequenos RNAs. 2. regulação epigenética. 3. estresses abióticos. 4. análise filogenética. 5. soja. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 570

EXPRESSÃO DE microRNAs NA GERMINAÇÃO DE *Glycine max* (L.)
MERRILL SOB ESTRESSES ABIÓTICOS E CARACTERIZAÇÃO DE UMA
FAMÍLIA GÊNICA ENVOLVIDA NA METILAÇÃO DO DNA DIRIGIDA POR
RNA EM PLANTAS

PAULA MACHADO DE ARAÚJO RIBEIRO

Tese apresentada ao Centro de Biociências e
Biotecnologia da Universidade Estadual do
Norte Fluminense Darcy Ribeiro, como parte
das exigências para a obtenção do título de
Doutora em Biociências e Biotecnologia.

Aprovada em 28 de fevereiro de 2024.

Banca examinadora:

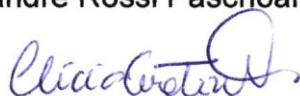


Prof.ª Antônia Elenir Amâncio Oliveira (D.Sc., Biociências e Biotecnologia) –
UENF



Prof.ª Ana Eliza Zeraik (D.Sc., Física Aplicada – Biomolecular) – UENF

Prof. Alexandre Rossi Paschoal (D.Sc., Bioinformática) – UTFPR



Prof.ª Clícia Grativol Gaspar de Matos (D.Sc., Química Biológica) – UENF
(Orientadora)



Governo do Estado do Rio de Janeiro
Universidade Estadual do Norte Fluminense Darcy Ribeiro
Diretoria do Centro de Biociências e Biotecnologia

DECLARAÇÃO

Declaro que a versão final da tese de Paula Machado de Araújo com título "Expressão de microRNAs na germinação de *Glycine max* (L.) Merrill sob estresses abióticos e caracterização de uma família gênica envolvida na metilação do DNA dirigida por RNA em plantas" foi aprovada pelo Prof. Alexandre Rossi Paschoal membro da banca impossibilitado de assinar a folha de aprovação.

Campos dos Goytacazes, 23 maio de 2024

Clicia Grativol Gaspar de Matos
Professora Associada
ID. Funcional 56492-8



Documento assinado eletronicamente por **Clicia Grativol Gaspar de Matos, Professora**, em 23/05/2024, às 09:09, conforme horário oficial de Brasília, com fundamento nos art. 28º e 29º do [Decreto nº 48.209, de 19 de setembro de 2022](#).



A autenticidade deste documento pode ser conferida no site http://sei.rj.gov.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=6, informando o código verificador **75227747** e o código CRC **8AC773B9**.

Referência: Processo nº SEI-260002/000310/2024

SEI nº 75227747

Avenida Alberto Lamego, 2000, - Bairro Pq. Califórnia, Campos dos Goytacazes/RJ, CEP 28013-602
Telefone: - www.uenf.br

*Dedico esta tese à minha mãe, Luza Maria:
Minha fonte de inspiração e apoio incondicional. Sua força,
amor e dedicação me impulsionaram a seguir em frente
nos momentos mais desafiadores. Esta conquista é tanto
minha quanto sua. Obrigada por sempre acreditar em mim,
mesmo quando eu duvidava de mim mesma.*

AGRADECIMENTOS

Nesse espaço, gostaria de registrar meus profundos agradecimentos a todas e todos que contribuíram para a concretização desta tese, em especial:

À minha orientadora, Dr.^a Clícia Grativol, por todos os ensinamentos, atenção e paciência ao longo desses anos de pesquisa, que se estendem desde 2016 quando ingressei na iniciação científica. Obrigada pelas diversas reuniões para esclarecer dúvidas e alinhar procedimentos, bem como pela disposição em ir para a bancada junto comigo quando estava aprendendo algumas das técnicas experimentais.

Aos meus pais, Luza Maria e Paulo, pelo suporte, por todos os esforços feitos que me permitiram chegar até aqui, e pela compreensão por eu precisar ficar longe.

Ao meu esposo, Elias, pela cumplicidade e por todo auxílio. Sou grata por cada gesto de carinho e por todas as palavras de incentivo que ajudaram a me fortalecer.

Aos amigos(as) do laboratório, Pedro, Felipe, Geovanna, Luciele, Giulia e Gabriela, por todos os momentos compartilhados e experiências trocadas.

Agradeço à Fernanda Coelho e Juliana Lopes por me auxiliarem durante os primeiros experimentos que realizei no doutorado e por me acompanharem no laboratório até altas horas, ainda durante a pandemia. À Geovanna Olimpio, pela ajuda prestada com algumas das técnicas de biologia molecular. À Pedro Igor pelo auxílio oferecido durante os experimentos com plântulas.

Aos coautores do artigo realizado durante o doutorado: Dr. Arthur Gruber (USP), Dr.^a Sara Sangi, Geovanna Olimpio e Felipe Cruz, pela colaboração com as análises realizadas.

Aos professores Dr.^a Antônia Elenir Amâncio, Dr.^a Ana Eliza Zeraik e Dr. Alexandre Paschoal por aceitarem compor a banca examinadora e por todas as contribuições.

À Dr^a Maria Aparecida Bertonceli, por todo apoio e pela prestatividade em aceitar revisar a tese.

À Marcela Rezende, pela amizade de anos e por todas as conversas de incentivo.

Aos laboratórios LQFPP e LBCT pela disponibilização dos equipamentos utilizados.

À Universidade Estadual do Norte Fluminense Darcy Ribeiro pela concessão da bolsa de doutorado, pela estrutura e por fornecer educação pública e de qualidade.

Ao Programa de Pós-graduação em Biociências e Biotecnologia por todo suporte.

A todos que, de forma direta ou indireta, fizeram parte dessa etapa da minha jornada acadêmica, muito obrigada!

“Digo: o real não está na saída nem na chegada: ele se dispõe para a gente é no meio da travessia.”

(Guimarães Rosa, em “Grande Sertão: Veredas”)

SUMÁRIO

RESUMO	xv
ABSTRACT	xvii
1. INTRODUÇÃO GERAL	1
1.1. Regulação epigenética em plantas.....	1
1.1.1. Pequenos RNAs	3
1.1.1.1. siRNAs.....	3
1.1.1.2. miRNAs.....	4
1.1.1.2.1. miPEPs.....	7
1.1.1.2.2. Regulação de miRNAs em condições de estresses abióticos	9
1.2. Aspectos gerais sobre a espécie <i>Glycine max</i> (soja).....	10
1.2.1. Estrutura e germinação de sementes de soja	11
2. OBJETIVOS	14
2.1. Objetivo geral	14
2.2. Objetivos específicos.....	14
3. CAPÍTULOS	15
3.1. Capítulo 1: Análise da expressão de microRNAs durante a germinação de sementes de <i>Glycine max</i> (L.) Merrill submetidas a estresses abióticos	15
3.1.1. RESUMO	15
3.1.2. INTRODUÇÃO.....	16
3.1.3. MATERIAIS E MÉTODOS	17
3.1.3.1. Germinação de sementes de soja sob condições de estresses abióticos	17
3.1.3.2. Avaliação fenotípica e teste estatístico	17
3.1.3.3. Análise dos perfis de expressão de miRNAs durante a germinação de soja após tratamento com manitol e NaCl.....	18
3.1.3.4. Predição de alvos do miR168	19
3.1.3.5. Aplicação de miRNA maduro durante a germinação	19
3.1.3.6. Predição e aplicação de miPEPs durante a germinação.....	19
3.1.3.7. Aplicação e avaliação da expressão do miR168 em plântulas de soja.	20
3.1.4. RESULTADOS.....	21
3.1.4.1. Estresse osmótico e salino afetam a germinação de soja	21
.....	25
3.1.4.2. Estresse osmótico e salino podem alterar os perfis de expressão de miRNAs durante a germinação de soja	25

3.1.4.3.	Efeitos do tratamento com miR168 exógeno	27
3.1.4.4.	Tratamento com miPEPs sintéticos associados ao miR168.....	29
3.1.4.5.	Aplicação do miR168 em plântulas de soja	30
3.1.5.	DISCUSSÃO	32
3.1.6.	CONCLUSÕES.....	36
3.1.7.	REFERÊNCIAS BIBLIOGRÁFICAS.....	37
3.2.	Capítulo 2: Identificação e evolução da família gênica CLASSY em genomas de plantas	40
3.2.1.	ABSTRACT	40
3.2.2.	INTRODUCTION	41
3.2.3.	MATERIALS AND METHODS	44
3.2.3.1.	Identification of CLSY gene family members.....	44
3.2.3.2.	Phylogenetic analysis.....	45
3.2.3.3.	Gene structure and domain prediction.....	45
3.2.3.4.	Non-synonymous and synonymous (Ka/Ks) analysis for duplicated pairs of CLASSY genes in soybean	46
3.2.3.5.	Expression analysis of CLSY genes in Arabidopsis and soybean tissues	46
3.2.3.6.	Expression analysis of a CLSY gene during soybean germination under stress conditions.....	46
3.2.4.	RESULTS AND DISCUSSION.....	47
3.2.4.1.	Construction of specific profile HMMs of CLSY proteins	47
3.2.4.2.	Identification, phylogenetic relationship, and structural analysis of CLSY family in plants.....	49
3.2.4.3.	Duplication events of the CLSY genes in soybean.....	58
3.2.4.4.	Tissue expression profile of CLSY genes in Arabidopsis and soybean	59
3.2.4.5.	The expression profile of CLSY and five other genes involved in epigenetic regulation can be modulated under abiotic stresses during soybean germination.....	62
3.2.5.	CONCLUSIONS	64
3.2.6.	REFERENCES	66
4.	CONSIDERAÇÕES FINAIS.....	74
5.	REFERÊNCIAS BIBLIOGRÁFICAS	74
	APÊNDICE	81
	ANEXO.....	99

LISTA DE FIGURAS

Introdução geral:

- Figura 1:** Representação esquemática da via RdDM canônica em plantas. Adaptado de Pikaard e Scheid (2014)..... 4
- Figura 2:** Biogênese dos miRNAs em plantas. Adaptado de Jones-Rhoades, Bartel e Bartel (2006). 6
- Figura 3:** Esquema da biossíntese de miPEPs a partir de ORFs presentes em pri-miRNAs. Adaptado de Erokhina et al. (2023)..... 7
- Figura 4:** Esquema da estrutura básica interna e externa de sementes de soja. Adaptado de Medic, Atkinson e Hurburgh (2014)..... 12
- Figura 5:** Eventos físicos e metabólicos que marcam a germinação (Fases I e II) e pós-germinação (Fase III). A curva representa a absorção de água ao longo do tempo. Adaptado de Nonogaki, Bassel e Bewley (2010). 13

Capítulo 1:

- Figura 1:** Taxa de germinação de sementes das cultivares BR-16 e Embrapa 48 não-tratadas e tratadas com manitol e NaCl nas concentrações de 100, 200 e 300 mM, em diferentes horas após a embebição. (A) Sementes da cultivar BR-16 tratadas com manitol. (B) Sementes da cultivar Embrapa 48 tratadas com manitol. (C) Sementes da cultivar BR-16 tratadas com NaCl. (D) Sementes da cultivar Embrapa 48 tratadas com NaCl..... 22
- Figura 2:** Massa fresca dos eixos embrionários no controle e nos tratamentos com manitol e NaCl, nas concentrações de 100, 200 e 300 mM, no ponto de 30 HAE. (A) Tratamento com manitol em ambas as cultivares. (B) Tratamento com NaCl em ambas as cultivares. As letras a, b, c e d indicam diferença significativa na massa fresca dos eixos embrionários controle e tratados, conforme teste de Tukey. 23
- Figura 3:** Massa seca dos eixos embrionários no controle e nos tratamentos com manitol e NaCl, nas concentrações de 100, 200 e 300 mM, no ponto de 30 HAE. (A) Tratamento com manitol em ambas as cultivares. (B) Tratamento com NaCl em ambas as cultivares. As letras a, b, c e d indicam diferença significativa na massa seca dos eixos embrionários controle e tratados, conforme teste de Tukey. 24
- Figura 4:** Medida da área dos eixos embrionários no controle e nos tratamentos com NaCl e manitol, nas concentrações de 100, 200 e 300 mM, no ponto de 30 HAE. (A) Tratamento com manitol em BR-16. (B) Tratamento com manitol em Embrapa 48. (C) Tratamento com NaCl em BR-16. (D) Tratamento com NaCl em Embrapa 48. As letras a, b, c e d indicam diferença significativa na área dos eixos embrionários controle e tratados, conforme teste de Tukey..... 25
- Figura 5:** Perfis de expressão de 15 miRNAs relacionados a estresses durante a germinação de sementes de soja, tratadas com manitol a 300 mM e NaCl a 100 mM. A linha tracejada representa a expressão dos miRNAs nas sementes controle. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$). 27
- Figura 6:** Perfil de expressão do miR168 em eixos embrionários da cultivar Embrapa 48, em 30 HAE, após aplicação de 1 µg e 5 µg do miR168 maduro. Os

asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$).	28
Figura 7: Análises fenotípicas após tratamentos com manitol a 300 mM, 5 µg do miR168, e 5 µg do miR168 + manitol a 300 mM em Embrapa 48. (A) Massa fresca dos eixos em 30 HAE. (B) Massa seca dos eixos em 30 HAE. (C) Área dos eixos em 30 HAE. As letras a, b, c e d indicam diferença significativa entre os tratamentos, conforme teste de Tukey.	28
Figura 8: Perfil de expressão do miR168 e seus alvos AGO1 (Glyma.16G217300 e Glyma.09G167100), em eixos embrionários da cultivar Embrapa 48, em 30 HAE, após tratamento com 5 µg do miR168 maduro e manitol a 300 mM. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$).	29
Figura 9: Análises fenotípicas após tratamento com miPEPs em Embrapa 48. (A) Massa fresca dos eixos em 30 HAE. (B) Massa seca dos eixos em 30 HAE. (C) Área dos eixos em 30 HAE.	30
Figura 10: Perfis de expressão do miR168 e seus alvos AGO1 (Glyma.16G217300 e Glyma.09G167100), após tratamento com miPEPs em Embrapa 48. A linha tracejada representa a expressão dos genes nas amostras controle.	30
Figura 11: Análise fenotípica de plântulas de sete dias das cultivares BR16 e Embrapa 48 após tratamento com miR168 maduro. (A) Comprimento, (B) Massa fresca e (C) Massa seca de plântulas controle e tratadas da cultivar BR16. (E) Comprimento, (F) Massa fresca e (G) Massa seca de plântulas controle e tratadas da cultivar Embrapa 48.....	31
Figura 12: Perfil de expressão do miR168 e dos genes-alvo AGO1 (Glyma.16G217300 e Glyma.09G167100) em raízes de BR16 (A) e Embrapa 48 (B) tratadas com miR168 exógeno. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$)......	32
Capítulo 2:	
Fig. 1. Workflow of the protein search approach using profile HMMs of SNF2 and Helicase C in Arabidopsis, followed by phylogenetic analysis.....	48
Fig. 2. Phylogenetic tree of proteins containing the SNF2 and Helicase C domains from 11 species, highlighting the monophyletic group containing the CLSY and DRD1 proteins. IQ-TREE software was used to construct the tree utilizing the maximum likelihood method. The values showed on the branches represent the bootstrap values from 1000 pseudoreplicates. Light blue, dark blue, and purple represent clades 1, 2, and 3, respectively. The single <i>P. patens</i> protein is shown in green. The complete tree can be found in the supplementary material (Fig. S6).....	51
Fig. 3. Gene structure of CLSYs, considering the clade division shown in the phylogenetic tree. Exons and introns are represented by blue boxes and black lines, respectively. The length of exons and introns is indicated in kilobases (kb).	54
Fig. 4. Protein domain architecture following the clade division from the phylogenetic analysis. The size of proteins can be estimated from the scale in	

number of amino acids (aa). The corresponding color of each domain is indicated in the legend.....	55
Fig. 5. Multiple sequence alignment of the SNF2 domain obtained in Jalview, highlighting the clades 1, 2, and 3. The clades are in the order in which they were arranged in the phylogenetic tree (Fig. S6). A color scale on amino acids was used to indicate sequence similarity.....	57
Fig. 6. Log2 expression (TPM) of nine soybean genes from clades 1, 2, and 3 under stress conditions. Green and orange colors indicate low and high expression, respectively. (a) soybean seedlings treated with 150 mM NaCl after six hours. (b) soybean roots subjected to drought. (c) soybean roots treated with 0.9% NaCl after 1, 2, 4, 24, and 48 hours. (d) soybean shoots subjected to water deficit after 0, 6, 12, and 24 hours.....	62
Fig. 7 Expression profiles of the AGO1, AGO4, DCL3, DRM2, Glyma.U027200, and ROS1 genes during the germination of soybean seeds treated with 300 mM mannitol and 100 mM NaCl. The dashed line represents the expression of genes in control seeds. Asterisks indicate significantly different expression values between the control and treatments, according to the t test ($P < 0.05$).	64
Fig. S1. Phylogenetic tree of 41 proteins containing the SNF2 domain. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.....	90
Fig. S2. Phylogenetic tree with the 123 proteins featuring the Helicase C domain. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.....	91
Fig. S3. Phylogenetic tree with the 35 proteins in common between the SNF2 and Helicase C domains. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.....	92
Fig. S4. Phylogenetic tree containing the 41 SNF2 domains, but excluding the Helicase C domain of the 35 proteins with both domains. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.....	92
Fig. S5. Phylogenetic tree comprising the 123 Helicase C domains, but excluding the SNF2 domain of the 35 proteins. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.....	93
Fig. S6. Complete phylogenetic tree of the 447 identified proteins, with emphasis on the monophyletic group containing proteins related to the CLSY and DRD1 families. Clades 1, 2, and 3 are shown in light blue, dark blue, and purple, respectively. The values indicated on the branches represent the bootstrap values from 1000 replicates.....	94
Fig. S7. Multiple sequence alignment of the Helicase C domain obtained in Jalview, highlighting the clades 1, 2, and 3. The clades are in the order in which they were arranged in the phylogenetic tree (Fig. S6). A color scale on amino acids was used to indicate sequence similarity.	95
Fig. S8. Tissues in which the CLSY1-4, AT2G16390/DRD1, and AT2G21450 genes are most expressed in Arabidopsis. The scale in the Shoot Apex column refers to the expression of genes in shoot apex, seed, and developing seed	

tissues. The scale in the carpel column refers to the expression of genes in carpel, young seeds, and flower tissues.....	96
Fig. S9. Expression profile of Arabidopsis genes CLSY1-4, AT2G16390/DRD1, and AT2G21450 under nine types of abiotic stresses in shoot and root tissues. (a) CLSY1; (b) CLSY2; (c) CLSY3; (d) CLSY4; (e) AT2G16390/DRD1; (f) AT2G21450.....	97
Fig. S10. Expression profiles of soybean genes from the three clades in 14 different tissues. (a) Clade 1 genes; (b) Clade 2 genes; (c) Clade 3 genes. No results were reported for the gene Glyma.02G261800.....	98
Fig. S11. Relative expression of Glyma.U027200 and Glyma.08G339900 genes in embryonic axes of cultivar BRS284, which refers to the same transcriptome used in the Soybean Expression Atlas database. The asterisk indicates significantly different expression values between genes according to the t-test ($P < 0.05$).....	98

LISTA DE TABELAS

Capítulo 1:

Tabela S1: Sequências de primers utilizados nas análises de RT-qPCR..... 82

Tabela S2: Genes-alvo preditos para o miR168..... 84

Capítulo 2:

Table 1. Gene identification and sequence characteristics of the 56 corresponding proteins identified from the profile HMMs. 52

Table 2. Paralog pairs genes, Ka/Ks ratio values, and duplication type of soybean genes. 59

Table S1. Primer sequences (forward/reverse) used in RT-qPCR analysis..... 89

ABREVIATURAS

µg	micrograma
µL	microlitro
aa	Aminoácidos
ABA	Ácido abscísico
AGO	Argonauta
ARF	<i>Auxin Response Factor</i>
CLSY	CLASSY
CMT3	CHROMOMETHYLASE 3
DCL	<i>Dicer-like</i>
DRD1	DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
dsRNA	<i>double-stranded RNA</i>
HAE	Horas após a embebição
HEN1	HUA ENHANCER 1
HMM	<i>Hidden Markov Models</i>
HYL1	HYPONASTIC LEAVES 1
MET1	DNA METHYLTRANSFERASE 1
miPEP	<i>microRNA-encoded peptide</i>
miRNA	microRNA
miRNA*	microRNA estrela
nt	Nucleotídeos
ORF	<i>Open Reading Frames</i>
Pol	RNA Polimerase
pre-miRNA	microRNA precursor
pri-miRNA	microRNA primário
RdDM	<i>RNA-directed DNA methylation</i>
RDR2	RNA-DEPENDENT RNA POLYMERASE 2
RISC	<i>RNA-Induced Silencing Complex</i>
ROS1	REPRESSOR OF SILENCING 1
RT-qPCR	<i>Reverse transcription-quantitative polymerase chain reaction</i>
SHH1	SAWADEE HOMEODOMAIN HOMOLOG 1

siRNA	Pequeno RNA de interferência
sRNA	Pequeno RNA
ssRNA	<i>single-stranded RNA</i>
TE	Elemento transponível

RESUMO

A regulação epigenética envolve modificações nos genomas dos organismos que não alteram a sequência de nucleotídeos do DNA, mas que podem ser herdadas. Um dos mecanismos epigenéticos compreende a ação de pequenos RNAs não-codificantes (sRNAs), que desempenham um papel fundamental na regulação da expressão gênica em diversos processos de desenvolvimento em plantas, assim como respondem a estresses ambientais. Os sRNAs podem ser divididos em duas categorias principais em plantas: microRNAs (miRNAs) e pequenos RNAs de interferência (siRNAs). No primeiro capítulo deste trabalho, buscamos investigar os perfis de expressão de miRNAs durante a germinação de soja submetida a estresses abióticos (salino e osmótico). A expressão de 15 miRNAs relacionados a estresses foi avaliada em duas cultivares de soja com diferentes tolerâncias à seca, a partir de RT-qPCR. Foi visto que tanto o estresse salino, quanto o osmótico são capazes de alterar a expressão de alguns miRNAs durante a germinação, e que os perfis de expressão variam entre cultivares contrastantes. Também foi analisado se a aplicação do miR168 exógeno em sementes da cultivar em que foi *down*-regulado, seria capaz de aumentar a sua expressão. As sementes tratadas com miR168 maduro apresentaram a expressão do miR168 significativamente maior que as sementes controle em 30 HAE. Além disso, a expressão do miR168 permaneceu aumentada em raízes de plântulas após sete dias de crescimento. No segundo capítulo, foi feita a identificação e análise evolutiva de proteínas da família CLASSY (CLSY), envolvidas em uma via de metilação do DNA que é mediada por siRNAs (via RdDM). Novos potenciais membros da família CLSY foram identificados em 11 espécies de plantas, incluindo uma briófitas, uma angiosperma basal, uma eudicotiledônea basal, quatro monocotiledôneas e quatro eudicotiledôneas. A identificação das proteínas CLSY envolveu a construção de HMMs de perfil (modelos ocultos de Markov) específicos para as proteínas CLSY, com base nos domínios dessa proteína (SNF2 e Helicase C) já caracterizados em *Arabidopsis*, bem como a construção de uma árvore filogenética utilizando o método de máxima verossimilhança. A análise de expressão de um gene CLSY durante a germinação de soja submetida a estresse salino e osmótico, mostrou pela primeira vez que esse gene pode ser responsivo a estresses abióticos. Os dados

obtidos nesse trabalho fornecem novas informações sobre a dinâmica de regulação dos miRNAs em situações de estresse em soja, e contribuem para a investigação dos genes CLSY em plantas, numa perspectiva evolutiva.

Palavras-chave: pequenos RNAs; regulação epigenética; estresses abióticos; análise filogenética; soja.

ABSTRACT

Epigenetic regulation involves modifications in the genomes of organisms that do not alter the nucleotide sequence of DNA, but which can be inherited. One of the epigenetic mechanisms comprises the action of small non-coding RNAs (sRNAs), which play a fundamental role in regulating gene expression in various developmental processes in plants, as well as responding to environmental stresses. sRNAs can be divided into two main categories in plants: microRNAs (miRNAs) and small interfering RNAs (siRNAs). In the first chapter of this work, we investigated the expression profiles of miRNAs during the germination of soybean subjected to abiotic stresses (saline and osmotic). The expression of 15 stress-related miRNAs was evaluated in two soybean cultivars with different drought tolerances, using RT-qPCR. It was found that both salt and osmotic stresses can alter the expression of some miRNAs during germination, and that expression profiles vary between contrasting cultivars. It was also analyzed whether the application of exogenous miR168 to seeds of the down-regulated cultivar could increase its expression. Seeds treated with mature miR168 showed significantly higher miR168 expression than control seeds at 30 HAI (hours after imbibition). Furthermore, miR168 expression remained increased in seedling roots after seven days of growth. In the second chapter, was performed the identification and evolutionary analysis of CLASSY (CLSY) family proteins – involved in a DNA methylation pathway mediated by siRNAs (RdDM pathway). Potential new members of the CLSY family were identified in 11 plant species, including one bryophyte, one basal angiosperm, one basal eudicot, four monocots and four eudicots. The identification of CLSY proteins involved the construction of profile HMMs (hidden Markov models) specific to CLSY proteins, based on the domains of this protein (SNF2 and Helicase C) already characterized in *Arabidopsis*, as well as the construction of a phylogenetic tree using the maximum likelihood method. Expression analysis of a CLSY gene during germination of soybean subjected to salt and osmotic stress showed for the first time that this gene can be responsive to abiotic stresses. The data obtained in this work provide new information on the regulation dynamics of miRNAs in stress situations in soybean, and contribute to the investigation of CLSY genes in plants from an evolutionary perspective.

Keywords: small RNAs; epigenetic regulation; abiotic stresses; phylogenetic analysis; soybean.

1. INTRODUÇÃO GERAL

1.1. Regulação epigenética em plantas

A epigenética consiste no estudo das mudanças nos padrões de expressão gênica, capazes de serem transmitidas por mitose e/ou meiose, sem alterar a sequência de DNA. Os mecanismos de regulação epigenética compreendem modificações covalentes no DNA e histonas, que interferem na atividade transcricional da cromatina (IWASAKI; PASZKOWSKI, 2014). Apesar de serem reversíveis e não alterarem o código genético em si, essas modificações podem mudar a forma como as sequências de DNA são lidas (KUMARI et al., 2022).

O termo "epigenética" (epi, que significa "acima/além" + genética), utilizado pela primeira vez em 1942 (WADDINGTON, 2012), diz respeito a uma camada de informação além da que é codificada pelo DNA. Dessa forma, embora as células somáticas de um organismo contenham o mesmo material genético, os padrões de expressão dos genes podem variar entre os diferentes tipos celulares (BRUKHIN; ALBERTINI, 2021).

Por serem organismos sésseis, as plantas muitas vezes necessitam lidar com condições ambientais desfavoráveis. Através dos mecanismos de regulação epigenética, é possível que os padrões de expressão gênica das plantas sejam ajustados, permitindo que se adaptem e reproduzam nas diversas condições que lhe são impostas (PIKAARD; SCHEID, 2014). Dado que as modificações epigenéticas podem gerar variações fenotípicas herdáveis através das gerações, considera-se que potencialmente desempenham um papel importante na evolução e ecologia das plantas (LUCIBELLI; VALOROSO; ACETO, 2022; MIRYEGANEH; SAZE, 2020; RAPP; WENDEL, 2005). Essas modificações ocorrem de forma transcricional ou pós-transcricional, e incluem os mecanismos de metilação do DNA, modificações de histonas e pequenos RNAs (PIKAARD; SCHEID, 2014).

A metilação do DNA é uma das marcas epigenéticas mais bem caracterizadas em plantas e animais (FENG; JACOBSEN, 2011). Consiste na adição de um grupo metil (-CH₃) no carbono 5' de um anel de citosina e, em plantas, pode ocorrer nos contextos CG, CHG e CHH (onde H representa os nucleotídeos A, C ou T) (LAW; JACOBSEN, 2010). A metilação é catalisada por enzimas denominadas DNA metiltransferases, podendo ocorrer em citosinas não-metiladas (metilação *de novo*)

ou em citosinas previamente metiladas, mantendo os padrões de metilação. Em plantas, esse processo é conduzido pelas seguintes enzimas: DNA METHYLTRANSFERASE 1 (MET1), que realiza a manutenção da metilação no contexto CG; CHROMOMETHYLASE 3 (CMT3), que mantém a metilação no contexto CHG; e DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), que é responsável pela metilação *de novo* no contexto CHH, e também nos contextos CG e CHG (LAW; JACOBSEN, 2010; PIKAARD; SCHEID, 2014). A DRM2 é direcionada para a sequência alvo por pequenos RNAs, através da via de metilação do DNA dirigida por RNA (RdDM) (HENDERSON; JACOBSEN, 2007).

Os níveis de metilação do DNA são controlados pelo processo de desmetilação, isto é, pela remoção de citosinas metiladas, que são substituídas por citosinas não-metiladas. A desmetilação ativa é realizada por enzimas desmetilases, que incluem REPRESSOR OF SILENCING 1 (ROS1), DEMETER (DME), DEMETER-LIKE 2 (DML2) e DEMETER-LIKE 3 (DML3) (LI; KUMAR; QIAN, 2018; ZHU, 2009). A expressão da ROS1 é aumentada a partir da metilação do DNA. Isso ocorre porque na região promotora do gene ROS1 há uma sequência de monitoramento de metilação do DNA (MEMS) de 39 pares de bases (pb), que é regulada pela via RdDM e por MET1. Dessa forma, é possível que a planta controle o nível de metilação do DNA e alcance um equilíbrio entre a metilação e desmetilação, por meio do ajuste fino da transcrição de ROS1 (LEI et al., 2015; WILLIAMS et al., 2015; ZHANG; LANG; ZHU, 2018).

Um outro importante mecanismo de regulação epigenética é a modificação de histonas, proteínas nucleares envolvidas no empacotamento do DNA. Plantas possuem diversas variantes de histonas que são modificadas pós-traducionalmente. Essas modificações podem levar à remodelação da cromatina, que altera a disponibilidade do DNA, regulando a expressão gênica (BRUKHIN; ALBERTINI, 2021; PIKAARD; SCHEID, 2014). Alguns dos processos aos quais as histonas podem ser submetidas são acetilação, metilação, fosforilação e ubiquitinação (IWASAKI; PASZKOWSKI, 2014). As modificações de histonas podem ter impacto em processos de desenvolvimento, como germinação e floração, assim como na resposta a estresses ambientais (ZHAO; ZHAN; JIANG, 2019).

A regulação epigenética também envolve a atuação de pequenos RNAs (sRNAs), que integram uma classe de RNAs não-codificadores com comprimento de 20 a 24 nucleotídeos (nt) aproximadamente (SIMON; MEYERS, 2011). Os sRNAs

são capazes de silenciar a expressão de genes nos mais variados processos biológicos em plantas, e são classificados com base na via de biogênese em: pequenos RNAs de interferência (siRNAs) e microRNAs (miRNAs) (KUMAR et al., 2018). Essas duas classes serão abordadas com mais detalhes nos tópicos a seguir.

1.1.1. Pequenos RNAs

1.1.1.1. siRNAs

Os sRNAs mais abundantes na maioria dos genomas de plantas são os siRNAs de 24 nt, que são componentes fundamentais na via RdDM (MATZKE; MOSHER, 2014). A RdDM possui um papel importante no silenciamento de transposons, vírus, transgenes ou genes repetitivos (PIKAARD; SCHEID, 2014). Na via RdDM canônica (Figura 1), inicialmente, a RNA Polimerase IV (Pol. IV) específica de plantas produz RNAs de fita simples (ssRNAs), que são convertidos em RNAs de fita dupla (dsRNAs) pela RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). Nessa etapa, é necessária a interação da Pol. IV com as proteínas CLASSY 1 (CLSY1) e SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1), que atuam no recrutamento da Pol. IV para a cromatina (GALLEGO-BARTOLOMÉ, 2020). Os dsRNAs resultantes são clivados pela DICER-LIKE 3 (DCL3) em siRNAs de 24 nt, que por sua vez são metilados pela HUA ENHANCER 1 (HEN1) e incorporados na ARGONAUTA 4 (AGO4). O complexo AGO-siRNA se liga a um transcrito complementar produzido pela RNA Polimerase V (Pol. V), e a DRM2 é recrutada para catalisar a metilação da citosina em todos os contextos de sequência. O complexo DDR, composto por DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) e RNA-DIRECTED DNA METHYLATION 1 (RDM1), facilita o acesso da Pol. V a seus locais-alvo (ERDMANN; PICARD, 2020; PIKAARD; SCHEID, 2014). A metilação no contexto CHH é considerada uma marca específica da via RdDM, apesar de poder estabelecer a metilação em todos os contextos. Nos contextos CG e CHG, a metilação pode ser mantida pela metilação de manutenção. Porém, no contexto CHH, a perpetuação da metilação depende da metilação *de novo* direcionada pela via RdDM (LAW; JACOBSEN, 2010).

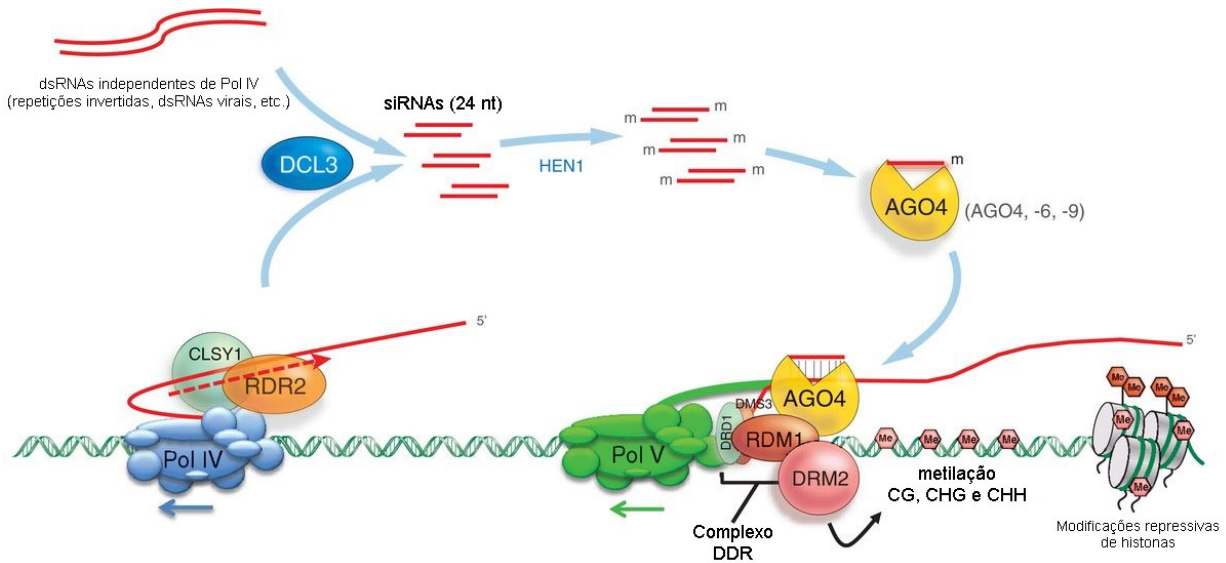


Figura 1: Representação esquemática da via RdDM canônica em plantas. Adaptado de Pikaard e Scheid (2014).

1.1.1.2. miRNAs

Os miRNAs são sRNAs com cerca de 21 nt de comprimento que desempenham papéis cruciais em diferentes tipos de organismos, incluindo protistas, fungos, plantas e animais (TAYLOR et al., 2014; VAUCHERET, 2006). A conservação de alguns miRNAs ao longo de longas distâncias evolutivas e a identificação de miRNAs na alga unicelular *Chlamydomonas reinhardtii* indicam que o mecanismo de regulação dos miRNAs é antigo e evolutivamente conservado (MOLNÁR et al., 2007; RAMACHANDRAN; CHEN, 2008). Em plantas, os miRNAs foram relatados pela primeira vez no ano de 2002, em *Arabidopsis thaliana* (LLAVE et al., 2002). Desde então, têm sido relacionados a diversos processos de desenvolvimento, como crescimento radicular lateral, manutenção de células meristemáticas, desenvolvimento da antera, vagem e semente, morfologia de folhas e pétalas, e promoção da floração (D'ARIO; GRIFFITHS-JONES; KIM, 2017; JONES-RHOADES; BARTEL; BARTEL, 2006).

A biogênese dos miRNAs se inicia com a transcrição de miRNAs primários (pri-miRNAs) pelo gene *MIR*, promovida pela RNA Polimerase II (Pol. II) (Figura 2). Os pri-miRNAs possuem estrutura de *hairpin*, com auto-complementaridade parcial, e são caracterizados por possuir um cap 7-metil-guanosina em sua extremidade 5' e uma cauda poli-A na extremidade 3' (LEE et al., 2004). A enzima DICER-LIKE 1 (DCL1), interagindo com a proteína nuclear HYPONASTIC LEAVES 1 (HYL1), converte o pri-

miRNA em um miRNA precursor (pre-miRNA). Em seguida, o pre-miRNA é processado pela DCL1 em um duplex contendo a fita guia ou miRNA maduro e a fita complementar, denominada miRNA* ou miRNA *star*. O duplex miRNA/miRNA* apresenta um fosfato (P) na posição 5' e dois nucleotídeos sobressalentes na posição 3'. O duplex miRNA/miRNA* é então metilado pela enzima metiltransferase nuclear HUA1 (HUA-ENHANCER 1) na extremidade 3' e transportado para o citoplasma pela proteína de membrana HASTY. A fita do miRNA maduro é incorporada ao complexo RISC (*RNA-Induced Silencing Complex*), que contém a proteína ARGONAUTA 1 (AGO1) como um componente central, enquanto o miRNA* é degradado. Junto ao complexo RISC, o miRNA irá atuar no silenciamento gênico (JONES-RHOADES; BARTEL; BARTEL, 2006; YU; JIA; CHEN, 2017).

O principal mecanismo de ação dos miRNAs em plantas é a clivagem do RNA mensageiro (mRNA) alvo, podendo também inibir a tradução (KIDNER; MARTIENSSEN, 2005). No mecanismo de clivagem do transcrito, o miRNA reconhece seu alvo a partir da complementaridade quase perfeita entre as sequências (YU; JIA; CHEN, 2017). Além disso, um estudo demonstrou que os miRNAs são capazes de regular a expressão de genes por meio da metilação do DNA. Wu e colaboradores (2010) descreveram uma via de miRNAs com 24 nt de comprimento, que requerem a DCL3 para clivar o precursor durante a biogênese, e são incorporados à AGO4. Esse mecanismo foi descrito em *Oryza sativa* (arroz), e mostrou que os miRNAs de 24 nt podem mediar a metilação do DNA nos próprios loci em que são produzidos, assim como nos seus genes-alvo (WU et al., 2010). Um outro estudo, em *Arabidopsis*, mostrou que a interação de miRNAs com as sequências codificantes dos genes PHABULOSA and PHAVOLUTA levou à metilação *downstream* do sítio de ligação dos miRNAs (BAO; LYE; BARTON, 2004). Esses dados indicam que além de clivarem ou reprimirem a tradução o mRNA alvo, alguns tipos de miRNAs podem atuar no silenciamento baseado na cromatina (MATZKE; MOSHER, 2014).

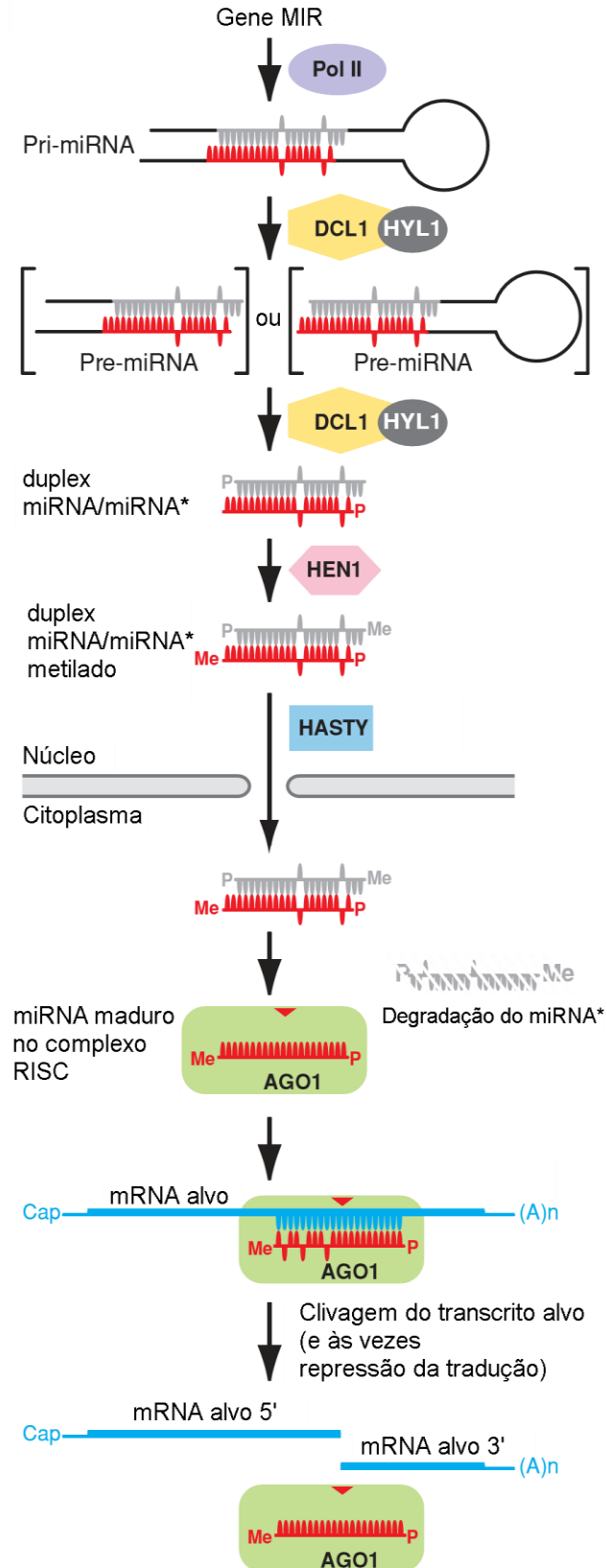


Figura 2: Biogênese dos miRNAs em plantas. Adaptado de Jones-Rhoades, Bartel e Bartel (2006).

1.1.1.2.1. miPEPs

No ano de 2015, foi relatado pela primeira vez que miRNAs primários em plantas são capazes de codificar pequenos peptídeos com função regulatória, denominados peptídeos codificados por microRNA (*microRNA-encoded peptides* – miPEPs) (Figura 3). Foram identificados miPEPs com tamanho de 3 a 59 resíduos de aminoácidos (aa), produzidos endogenamente a partir de quadros abertos de leitura (*Open Reading Frames* - ORFs) contidos nos pri-miRNAs. Os miPEPs foram capazes de aumentar especificamente a transcrição de seu pri-miRNA associado, consequentemente levando ao acúmulo do miRNA maduro. O pri-miR165a de *Arabidopsis* e o pri-miR171b de *Medicago truncatula* produziram, respectivamente, os miPEP165a e miPEP171b, que aumentaram a síntese de seus miRNAs maduros, o que levou a uma regulação negativa dos genes-alvo relacionados com o desenvolvimento de raízes. miPEPs ativos foram codificados por outros cinco pri-miRNAs de *Arabidopsis* e *M. truncatula*, indicando uma possível ubiquidade desses peptídeos no reino vegetal (LAURESSERGUES et al., 2015).

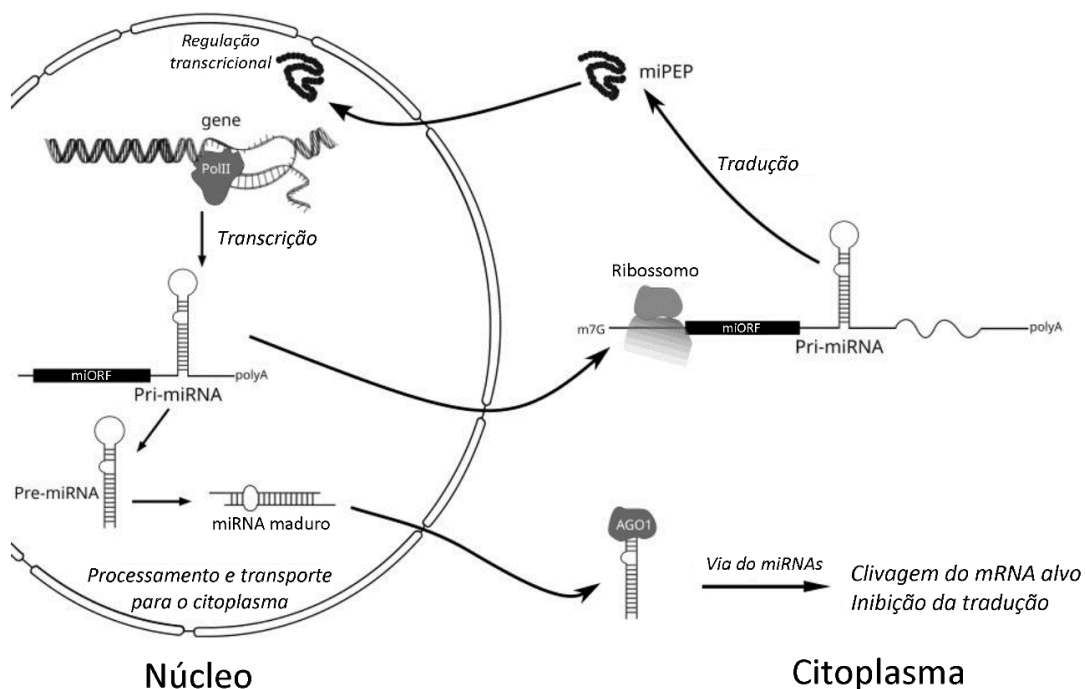


Figura 3: Esquema da biossíntese de miPEPs a partir de ORFs presentes em pri-miRNAs. Adaptado de Erokhina et al. (2023).

Desde a descoberta dos miPEPs há nove anos, diferentes estudos identificaram miPEPs em outras espécies de plantas, como: *Glycine max* (soja)

(COUZIGOU et al., 2016), *Vitis vinifera* (videira) (CHEN et al., 2020), e em espécies da família Brassicaceae (EROKHINA et al., 2021). miPEPs preditos computacionalmente também foram relatados em *Arachis hypogaea* (amendoim) (RAM; MUKHERJEE; PANDEY, 2019), *Phaseolus vulgaris* (feijão comum) e *Vigna unguiculata* (feijão-de-corda) (ARAÚJO; GRATIVOL, 2022). Alguns estudos realizaram tratamento exógeno com miPEPs, e demonstram que a aplicação de miPEPs sintéticos gera diferentes efeitos fenotípicos em plantas, a partir do aumento da expressão dos seus miRNAs associados (ORMANCEY et al., 2023). Por exemplo, em soja, a aplicação exógena do miPEP172c aumentou a abundância do miR172c e estimulou a formação de nódulos radiculares, indicando que o tratamento com esses peptídeos pode ter impacto tanto no desenvolvimento da planta, como na interação planta-microrganismo (COUZIGOU et al., 2016). Em *V. vinifera*, a aplicação do miPEP171d1 sintético foi capaz de promover o desenvolvimento de raízes adventícias, ao aumentar especificamente a expressão do miR171d (CHEN et al., 2020). Em um segundo trabalho realizado pelo mesmo grupo, plântulas de *V. vinifera* foram tratadas com outros dois miPEPs exógenos (miPEP172b e miPEP3635b), e apresentaram maior tolerância ao frio em relação às plântulas controle. Pela primeira vez, foi relatado que miPEPs derivados de pri-miRNAs podem atuar na resistência das plantas a um estresse ambiental (CHEN et al., 2022).

De forma geral, as sequências de miPEPs são pouco conservadas entre as espécies, ao contrário dos miRNAs, que costumam ser altamente conservados. Essa característica pode ser atribuída ao fato de que a atividade do miPEP depende diretamente da sequência de nucleotídeos que o codifica. Dessa forma, a sequência codificante da própria ORF é fundamental para a especificidade do miPEP, elucidando como este exerce funções específicas apenas na espécie hospedeira (LAURESSERGUES et al., 2022; ORMANCEY et al., 2023).

Em virtude do feedback positivo que os miPEPs são capazes de exercer sobre seus pri-miRNAs, podem desempenhar diversas funções biológicas em plantas, sendo envolvidos em processos de desenvolvimento, interações entre a planta e microrganismos e tolerância a estresse (ORMANCEY et al., 2023). Por meio da aplicação de miPEPs de forma exógena, mudanças fenotípicas podem ser induzidas. Dessa forma, os miPEPs surgem como uma possível ferramenta alternativa para a regulação de genes sem a necessidade de manipulações genômicas. Além disso,

podem ser uma opção aos pesticidas e produtos químicos nocivos ao ambiente (ORMANCEY et al., 2021; YADAV et al., 2021).

1.1.1.2.2. Regulação de miRNAs em condições de estresses abióticos

Projeções sobre as alterações no clima indicam impactos na segurança alimentar e, particularmente, no rendimento e qualidade da produção agrícola (FAROOQ et al., 2022). Em escala global, a mudança climática interanual é responsável por gerar cerca de um terço das variações nos rendimentos de culturas como soja, milho, trigo e arroz (RAY et al., 2015). Novas ferramentas biotecnológicas têm sido investigadas, visando culturas resilientes a alterações climáticas, e levando em conta o desenvolvimento sustentável. Dentre essas ferramentas, os miRNAs vêm sendo alvos de diferentes estudos por regularem a expressão gênica de plantas em condições de estresses ambientais (PATIL et al., 2021; RAZA et al., 2023; ZHANG; WANG, 2016).

Os miRNAs modulam respostas a diferentes tipos de estresses abióticos em plantas, como estresse salino, seca, calor, frio, hipóxia, radiação UV-B, estresse oxidativo e por metais pesados (KHRAIWESH; ZHU; ZHU, 2012; SUNKAR; LI; JAGADEESWARAN, 2012). Essas moléculas regulatórias também possuem um papel importante na recuperação de estresses, quando, após a exposição a um estresse, a planta necessita alcançar uma nova homeostase (CRISP et al., 2016). Um estudo feito por Stief et al. (2014) evidenciou que o miR156 está relacionado com a termotolerância adquirida em *Arabidopsis*. O miR156 tem como alvos os fatores de transcrição SQUAMOSA-PROMOTER BINDING-LIKE (SPL), reguladores-chave de transições de desenvolvimento. Os mutantes para a AGO1, envolvida com a biogênese do miRNA, tiveram a termotolerância prejudicada. Além disso, após a exposição das plantas ao estresse térmico, a indução do miR156 durante o período de recuperação do estresse foi crucial para aumentar a memória de estresse térmico (STIEF et al., 2014).

A seca e a salinidade são duas das principais limitações para a produção agrícola em todo o mundo (SUNKAR; LI; JAGADEESWARAN, 2012). Diferentes famílias de miRNAs já foram relatadas como responsivas a esses estresses. Por exemplo, em *Arabidopsis*, miR168, miR171 e miR396 respondem tanto à seca (tratamento com manitol), quanto à salinidade (tratamento com NaCl), bem como ao estresse por frio (4°C). Portanto, foi visto que um único miRNA tem potencial para regular diferentes genes-alvo relacionados a estresses (LIU et al., 2008).

Os miRNAs, assim como siRNAs e mRNAs possuem uma importante habilidade de se moverem por distâncias curtas ou longas através dos tecidos vegetais, transportando-se, por exemplo, de uma célula para outra por plasmodesmas ou entre diferentes tecidos via floema (MAIZEL et al., 2020). Um estudo em *Brassica napus* demonstrou que os miRNAs miR395, miR398 e miR399, identificados na seiva do floema, apresentaram níveis aumentados em resposta ao estresse por privação dos nutrientes sulfato, cobre e fosfato, respectivamente. As respostas no floema demonstram que os miRNAs desempenham um papel na transferência sistêmica de informações, por meio do transporte de longa distância, e podem indicar quando há deficiência de nutrientes na planta (BUHTZ et al., 2008).

Ademais, vale ainda ressaltar as potenciais aplicações dos miRNAs como marcadores moleculares sensíveis a estresses ambientais. A obtenção do perfil de expressão de miRNAs em plantas submetidas a estresses pode auxiliar na identificação de novos biomarcadores, visando o desenvolvimento de estratégias que contribuam para a tolerância a estresses em plantas (BEJ; BASAK, 2014; RAŽNÁ; ŽIAROVSKÁ; GÁLOVÁ, 2019).

1.2. Aspectos gerais sobre a espécie *Glycine max* (soja)

A soja é pertencente à família Fabaceae, gênero *Glycine* e espécie *Glycine max* (L.) Merrill. Foi domesticada a partir da espécie selvagem *Glycine soja* na China, há cerca de 5000 anos (GAO et al., 2018; HYMOWITZ, 1970). A soja é uma importante fonte de proteínas e óleo para a alimentação, além de também ser utilizada para a produção de biocombustíveis (XU et al., 2020). Adicionalmente, através da relação de simbiose com bactérias fixadoras de nitrogênio, a soja contribui para a melhoria da qualidade do solo (ANDERSON et al., 2019).

No Brasil, a soja foi introduzida no ano de 1882 em Salvador/Bahia e, mais adiante, na região Sul do país. Uma expansão significativa da cultura ocorreu na década de 1960, e a produtividade média aumentou através dos anos. Na década de 1980, a soja se expandiu para o Centro-Oeste e, aos poucos, se estabeleceu nas diferentes regiões do país (CATTELAN; DALL'AGNOL, 2018).

Entre as leguminosas, a soja foi a primeira espécie a ter seu genoma sequenciado, no ano de 2010. Com o total de 20 pares de cromossomos ($2n=40$) e um genoma de 1115 megabases (Mb), foram previstos 46.430 genes codificadores de

proteínas. *G. max* é considerada um antigo poliploide, em consequência das duplicações que ocorreram há cerca de 59 e 13 milhões de anos. Por esse motivo, quase 75% dos genes ocorrem em cópias múltiplas (SCHMUTZ et al., 2010). A soja constitui uma importante planta modelo de leguminosas. Os avanços nas últimas décadas na área de genômica da soja proporcionaram uma vasta quantidade de dados úteis para a realização de pesquisas básicas e aplicadas envolvendo essa espécie (DHAUBHADEL; MARSOLAIS, 2012).

Atualmente, a soja é a segunda maior *commodity* produzida no Brasil, depois da cana-de-açúcar (FAO, 2022). O Brasil ocupa a primeira posição entre os maiores produtores mundiais de soja, seguido pelos Estados Unidos. No ano de 2023, o Brasil produziu 157 milhões de toneladas do grão, correspondendo a aproximadamente 39% da produção mundial (USDA, 2023). Na safra de 2023/24, a produtividade média esperada no Brasil foi reduzida, devido às influências de extremos climáticos em algumas partes do país, como excesso de precipitações na região Sul e escassez de chuva nas demais regiões, aliada às altas temperaturas (CONAB, 2024). Juntos, esses dados revelam a importância que a cultura de soja possui no país e no mundo, e reforçam a necessidade do desenvolvimento de tecnologias visando mitigar os efeitos de variações climáticas extremas.

1.2.1. Estrutura e germinação de sementes de soja

A semente de soja é basicamente constituída por duas partes principais: o embrião e o tegumento ou testa (SUN; YUAN, 2022) (Figura 4). O embrião corresponde a 90% do peso da semente, sendo composto por dois cotilédones e o eixo embrionário. Os cotilédones funcionam como um tecido de armazenamento, com reservas fundamentais para o processo germinativo e desenvolvimento inicial. O eixo embrionário, por sua vez, se divide em radícula, hipocótilo e epicótilo. A radícula dará origem à raiz principal, o hipocótilo eleva os cotilédones e o epicótilo acima da superfície do solo, e o epicótilo corresponde à região do caule da futura planta (MEDIC; ATKINSON; HURBURGH, 2014). O tegumento constitui a barreira protetora entre o embrião e o ambiente externo. Já o hilo é uma "cicatriz" que o tegumento apresenta ao se separar da planta-mãe. É o local de conexão entre a semente e a vagem (BEWLEY et al., 2013; SUN; YUAN, 2022).

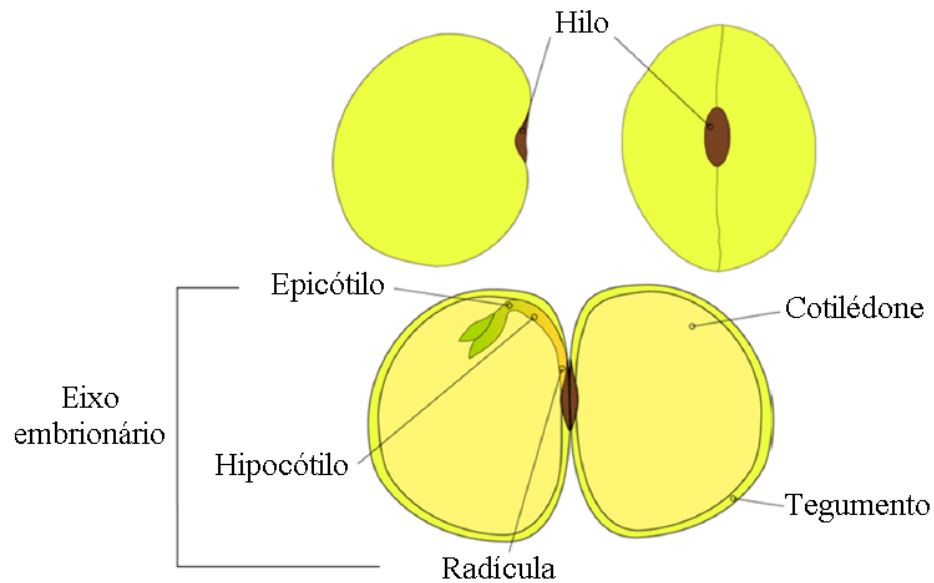


Figura 4: Esquema da estrutura básica interna e externa de sementes de soja. Adaptado de Medic, Atkinson e Hurburgh (2014).

A germinação de sementes consiste no processo fisiológico que culmina com a emergência do eixo embrionário pelo tegumento. Esse processo pode ser dividido em três fases correspondentes às fases de absorção da água (Figura 5). Na fase I, sementes maduras e secas (quiescentes) absorvem água (embebição) até que fiquem totalmente hidratadas. A partir da absorção de água, atividades metabólicas se iniciam, como respiração, reparo mitocondrial e multiplicação, reparo de DNA, transcrição e tradução de novos mRNAs, e tradução de mRNAs armazenados (BEWLEY et al., 2013). Na segunda fase, a embebição de água se estabiliza e os processos metabólicos aumentam. Nessa fase, ocorre a emergência da radícula devido à expansão celular, e o tegumento é rompido, o que caracteriza o fim da germinação (LUJÁN-SOTO; DINKOVA, 2021). Na fase III, correspondente à pós-germinação, a absorção de água aumenta e a plântula inicia o crescimento, utilizando as principais reservas armazenadas (NONOGAKI; BASSEL; BEWLEY, 2010).

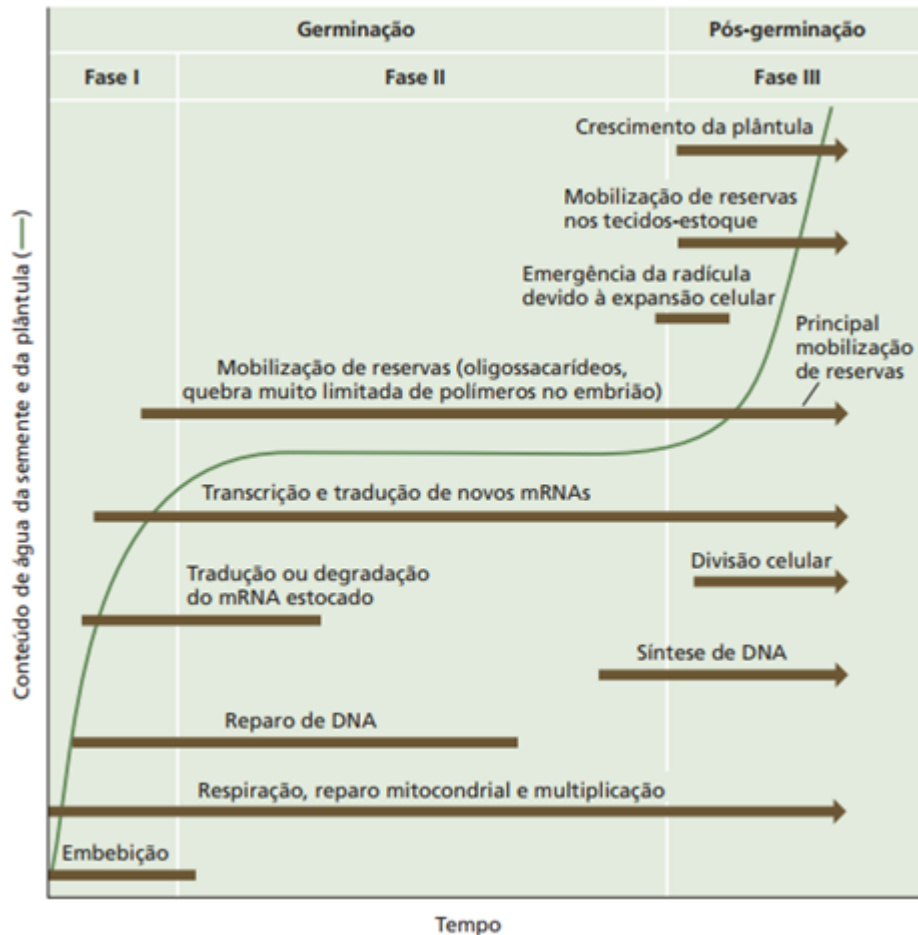


Figura 5: Eventos físicos e metabólicos que marcam a germinação (Fases I e II) e pós-germinação (Fase III). A curva representa a absorção de água ao longo do tempo. Adaptado de Nonogaki, Bassel e Bewley (2010).

A germinação pode ser regulada de diferentes formas, além de sofrer influências de fatores ambientais, como temperatura, luz, disponibilidade de água e fertilidade do solo (BEWLEY et al., 2013). Dentre os mecanismos de regulação da germinação mais conhecidos, estão os processos mediados por fitormônios. Por exemplo, as giberelinas (GAs) promovem a liberação da dormência e germinação das sementes, favorecem a expansão celular e o crescimento da plântula. Enquanto o ácido abscísico (ABA) atua como inibidor da germinação, sendo que esse hormônio em baixos níveis é um requisito obrigatório para a semente germinar. Dessa forma, o balanço entre ABA-GA é uma característica reguladora central da germinação (BEWLEY et al., 2013; LUJÁN-SOTO; DINKOVA, 2021). Outros hormônios também participam da regulação desse processo. Por exemplo, o etileno (ET), que possui um papel importante na liberação da dormência e, portanto, antagoniza os efeitos do ABA (LUJÁN-SOTO; DINKOVA, 2021). E a auxina (AUX), que pode atuar na dormência de

maneira dependente de ABA, alterando o conteúdo ou a sinalização de ABA (CARRERA-CASTAÑO et al., 2020).

2. OBJETIVOS

2.1. Objetivo geral

Analisar os perfis de expressão de miRNAs durante a germinação de sementes de *G. max* em condições de estresses abióticos, e identificar e analisar evolutivamente proteínas da família CLSY, envolvidas na via RdDM, em diferentes espécies de plantas.

2.2. Objetivos específicos

- Avaliar os efeitos do estresse salino e osmótico nas características fenotípicas de sementes de soja durante a germinação, considerando uma cultivar tolerante à seca e uma suscetível;
- Analisar os perfis de expressão de miRNAs em eixos embrionários das sementes submetidas à estresses;
- Avaliar a viabilidade do tratamento exógeno com miRNA maduro e miPEPs em sementes de soja;
- Identificar novos membros da família CLSY em diferentes espécies de plantas;
- Analisar a filogenia e conservação das proteínas CLSY identificadas;
- Avaliar os efeitos do estresse salino e osmótico na expressão de um gene CLSY e outros cinco genes envolvidos na regulação epigenética, em sementes de soja.

Nota: os procedimentos metodológicos utilizados neste trabalho, assim como os resultados obtidos e sua discussão foram organizados em dois capítulos, sendo o capítulo 1 escrito em formato convencional e o capítulo 2 em formato de artigo, a ser submetido na revista internacional *Plant Physiology and Biochemistry* (editora Elsevier).

3. CAPÍTULOS

3.1. Capítulo 1: Análise da expressão de microRNAs durante a germinação de sementes de *Glycine max* (L.) Merrill submetidas a estresses abióticos

3.1.1. RESUMO

A germinação de sementes se caracteriza pela protusão do eixo embrionário através do tegumento. A regulação desse processo é crucial para a produção de grãos de relevante valor nutricional e econômico, assim como para a multiplicação de espermatófitas. A germinação pode ser influenciada por condições ambientais, como disponibilidade de água e salinidade do solo. MicroRNAs (miRNAs) são RNAs não-codificadores que regulam a expressão gênica, podendo clivar RNAs mensageiros-alvo e inibir a tradução. A desregulação de miRNAs pode comprometer o processo germinativo e, conseqüentemente, o crescimento da plântula na pós-germinação. Esse trabalho visou analisar o perfil de expressão de miRNAs envolvidos com a germinação de sementes de soja (*Glycine max*) em condições de estresse osmótico e salino. Inicialmente, as sementes das cultivares BR-16 (sensível à seca) e Embrapa 48 (tolerante à seca) foram desinfestadas e tratadas com NaCl e manitol nas concentrações de 100 mM e 300 mM, respectivamente. As sementes tratadas e não-tratadas foram germinadas em B.O.D. a 28 °C, sem fotoperíodo. Com 30 horas após a embebição (HAE), os eixos embrionários foram seccionados manualmente para extração do RNA total e síntese de cDNA. Os perfis de expressão de 15 miRNAs relacionados com estresses abióticos foram avaliados por RT-qPCR, utilizando o equipamento Step One (Applied Biosystems). Os dados obtidos foram submetidos ao Teste t com significância de 95% ($P < 0,05$), no programa GraphPad Prism. Dentre os miRNAs com diferença significativa de expressão entre os tratamentos e o controle, o miR168 apresentou um nível de expressão significativamente maior em eixos embrionários da cultivar BR-16 tratados com NaCl e manitol. Já na cultivar Embrapa 48, o miR168 exibiu expressão significativamente menor nos dois tratamentos. Após, foram aplicados 1 µg e 5 µg do miR168 maduro em sementes da cultivar Embrapa 48 em 0 HAE. Com 30 HAE, foi feita extração e quantificação de RNA, RT-qPCR, e teste estatístico. As sementes tratadas com miR168 maduro apresentaram a expressão do miR168 significativamente maior que as sementes controle. Verificamos também que, em raízes de plântulas de sete dias, o miR168 aplicado antes da germinação

permanece com alta expressão e pode alterar a expressão de genes-alvo. Os resultados dessa pesquisa podem ajudar a elucidar o papel do miRNAs durante a germinação da soja submetida a estresses abióticos. A obtenção do perfil de expressão de miRNAs em condições de estresse pode auxiliar no desenvolvimento de estratégias que contribuam para a tolerância a estresses abióticos em plantas.

Palavras-chave: germinação; miRNAs; estresse salino; estresse osmótico; soja.

3.1.2. INTRODUÇÃO

É sabido que a germinação de sementes é finamente regulada pela ação de fitormônios. Além disso, as modificações epigenéticas e os pequenos RNAs também vêm sendo descritos como participantes fundamentais nesse processo (LUJÁN-SOTO; DINKOVA, 2021). Mais especificamente, diferentes estudos evidenciaram as funções que os miRNAs desempenham durante a germinação de sementes. Essas moléculas regulam principalmente fatores de transcrição e proteínas envolvidas em diversas vias de sinalização (DAS et al., 2015). Por exemplo, a família miR159 regula fatores de transcrição da família MYB (MYB33 e MYB101), que são reguladores positivos de ABA durante a dormência e germinação de sementes em *Arabidopsis* (REYES; CHUA, 2007). Outros tipos de fatores de transcrição regulados por miRNAs são os ARF (fator de resposta a auxina), das famílias ARF10, ARF16, ARF17, regulados pelo miR160. Num estudo feito com ARF10, também em *Arabidopsis*, foi visto que a regulação negativa de ARF10 pelo miR160 é fundamental para manter a comunicação-cruzada normal de auxina-ABA durante a germinação de sementes e desenvolvimento de plântulas. Foi demonstrado que a perda de função em miRNAs resulta em letalidade embrionária, defeitos embrionários graves ou formação anormal de plântulas após a germinação (LIU et al., 2007). Além disso, pesquisas apontam que diferentes famílias de miRNAs estão envolvidas com a resposta a estresses ambientais durante a germinação em *Arabidopsis*. A família miR402 regula a expressão da proteína DEMETER-LIKE 3, sendo demonstrada a função desse miRNA como regulador positivo da germinação de sementes e crescimento de plântulas de *Arabidopsis*, sob condições de estresse salino, seca e frio (KIM et al., 2010a). A família miR395 também está relacionada com a regulação da germinação sob estresse hídrico e salino, regulando a expressão da proteína ATP sulfúrilase (APS) e do

transportador de sulfato SULTR2. A superexpressão da isoforma miR395e acelerou a germinação de sementes de *Arabidopsis* mediante desidratação e a superexpressão de miR395c retardou a germinação sob condições de estresse salino (KIM et al., 2010b). De forma semelhante, a família miR417 exibe regulação negativa da germinação sob estresse salino (JUNG; KANG, 2007).

Apesar do importante papel dos miRNAs na modulação de estresses durante o processo germinativo, os relatos na literatura científica sobre a atuação desses pequenos RNAs na germinação de sementes de soja em condições de estresses ambientais ainda são escassos.

3.1.3. MATERIAIS E MÉTODOS

3.1.3.1. Germinação de sementes de soja sob condições de estresses abióticos

As sementes de soja das cultivares BR-16 (sensível à seca) e Embrapa 48 (tolerante à seca) foram esterilizadas com solução de hipoclorito de sódio a 2% por 30 segundos e lavadas com água destilada estéril cinco vezes. A germinação das sementes foi realizada em placas de Petri (90 × 15 mm) com 12 ml de água destilada esterilizada em papel germitest. As sementes submetidas a estresse osmótico foram tratadas com manitol, nas concentrações de 0 (controle), 100, 200 e 300 mM. As sementes expostas ao estresse salino foram tratadas com cloreto de sódio (NaCl), também nas concentrações de 0 (controle), 100, 200 e 300 mM. Para cada tratamento foram utilizadas duas réplicas biológicas, sendo cada réplica composta por um *pool* de 10 sementes. As placas de Petri foram mantidas em incubadora B.O.D. (Biochemical Oxygen Demand) sem fotoperíodo, com temperatura de 28°C, por 30 horas. A taxa de germinação foi avaliada nos pontos de 12, 15, 18, 21, 24, 27 e 30 horas.

3.1.3.2. Avaliação fenotípica e teste estatístico

Ao completar 30 HAE (horas após a embebição), os eixos embrionários foram seccionados manualmente para mensuração da área e análise de biomassa. A área dos eixos nos diferentes tempos e tratamentos foi mensurada com o uso do programa ImageJ (<https://imagej.nih.gov/ij/>) (ABRÀMOFF; MAGALHÃES; RAM, 2004). Foi

realizada a secagem dos eixos embrionários em estufa a 60°C por 24 h. Para a análise de biomassa, os eixos frescos e secos foram pesados para estimar, respectivamente, a massa fresca e seca. Os dados de área e biomassa obtidos foram submetidos ao teste de Tukey, com nível de significância de 95%. A análise estatística foi realizada no programa GraphPad Prism, versão 9.0 (<https://www.graphpad.com/>).

3.1.3.3. Análise dos perfis de expressão de miRNAs durante a germinação de soja após tratamento com manitol e NaCl

Foi feita uma busca na literatura por miRNAs relacionados a estresses abióticos em plantas (BARRERA-FIGUEROA et al., 2011; KULCHESKI et al., 2011; LI et al., 2011; LIU et al., 2008; ZHENG et al., 2016), e foram selecionados 15 miRNAs para análise da expressão nas condições de estresse aplicadas em soja: miR156, miR159, miR160, miR166, miR167, miR168, miR171, miR319, miR390, miR394, miR398, miR408, miR482, miR1507 e miR3522. A expressão dos miRNAs foi analisada por RT-qPCR. Para isso, primeiramente, foi realizada a extração de RNA total dos eixos embrionários com 30 HAE, tratados e não-tratados com 100 mM de NaCl e 300 mM de manitol, utilizando o reagente Trizol® (Invitrogen), conforme as instruções do fabricante. A quantificação de RNA foi realizada em espectrofotômetro NanoDrop One™ (Thermo Fisher Scientific). A integridade do RNA foi verificada por eletroforese em gel de agarose a 1%, corado com brometo de etídio. A síntese de cDNA foi feita com 5 µg de RNA, utilizando o kit SuperScript™ III (Invitrogen), de acordo com o protocolo do fabricante. O RT-qPCR foi realizado utilizando o StepOne™ Real-Time PCR System (Applied Biosystems), em placas de 48 poços com volume final de 10 µL, contendo 0,5 µL de cada primer direto e reverso, 5 µL de SYBR Green, 2 µL de cDNA e 2,5 µL de água ultrapura. Os primers de miRNAs foram construídos conforme descrito por Varkonyi-Gasic e Hellens (2011). O gene *S-adenosylmethionine synthase 4* (METK4) foi utilizado como constitutivo. Também foi analisada a expressão de dois genes AGO1 (Glyma.16G217300 e Glyma.09G167100), como alvos do miR168. Os primers de METK4 e AGO1 foram construídos usando a ferramenta *Oligo Analyzer* da *Integrated DNA Technologies* (IDT) (<https://www.idtdna.com/>). As sequências de todos os primers utilizados se encontram na Tabela S1. A expressão relativa foi calculada usando o método $2^{(-\Delta\Delta CT)}$ (RAO et al., 2013). Os dados obtidos foram submetidos ao Teste t com nível de significância de 95% ($P < 0,05$), no programa GraphPad Prism v.9.0 (<https://www.graphpad.com/>).

3.1.3.4. Predição de alvos do miR168

Os genes-alvo do miR168 foram preditos no programa psRNATarget, utilizando a versão do genoma de soja Wm82.a2.v1, com parâmetros padrão (<https://www.zhaolab.org/psRNATarget/home>) (DAI; ZHUANG; ZHAO, 2018).

3.1.3.5. Aplicação de miRNA maduro durante a germinação

Para a realização desse experimento, foi feita a síntese da fita do miR168 maduro (TCGCTTGGTGCAGGTCGGGAA) na Exxtend (<https://www.exxtend.com.br/>), sendo purificado por dessalinização e ressuspendido em tampão TE (pH 8,0). Para o tratamento exógeno com esse miRNA, inicialmente, as sementes da cultivar Embrapa 48 foram esterilizadas como descrito no tópico 3.1.3.1, e colocadas em placas para cultivo de células de 12 poços, sendo utilizadas duas sementes em cada poço. As sementes foram tratadas com 1 µg ou 5 µg do miR168 maduro (diluído em 50 µL) em cada poço, havendo seis poços (12 sementes) para cada tratamento. No tratamento controle, foram adicionados 50 µL de água Milli-Q autoclavada em cada poço. Após 3 horas de embebição, em todos os poços foram adicionados 600 µL de água Milli-Q autoclavada. As sementes foram germinadas em incubadora B.O.D. sem fotoperíodo, com temperatura de 28°C, até o ponto de 30 horas. Em 30 HAE, os eixos embrionários foram seccionados manualmente para a extração de RNA total, síntese de cDNA e RT-qPCR, seguindo o mesmo procedimento delineado no tópico 3.1.3.3. Num experimento seguinte, foi feita a aplicação do miR168 e adição do manitol. Foram utilizadas seis placas de cultivo de células, com os seguintes tratamentos em cada: controle, manitol a 300 mM, 5 µg do miR168, e 5 µg do miR168 + manitol a 300 mM. Dessa forma, para cada tratamento foram feitas seis réplicas com um *pool* de seis sementes em cada réplica. Foram seguidos os mesmos procedimentos de esterilização das sementes, condições de germinação e RT-qPCR apresentados anteriormente. Em 30 HAE, os eixos embrionários foram removidos para análise fenotípica, como descrito no tópico 3.1.3.2.

3.1.3.6. Predição e aplicação de miPEPs durante a germinação

Foi feita a predição de miPEPs relacionados ao miR168 seguindo o método descrito por Araújo e Grativol (2022). Foram selecionados dois miPEPs putativos para a síntese: MRLRWGF (miPEP1), com 7 aa, e MPCDCVWGF (miPEP2), com 9 aa. Os peptídeos foram sintetizados por Aminotech (<https://www.aminotech.net.br/>), com 95% de pureza, sendo feita análise de espectrometria de massas para comprovar a identidade da molécula e análise de HPLC para comprovar o grau de pureza. Ambos os miPEPs foram ressuspensos em água ultrapura. O tratamento com os miPEPs foi feito em sementes da cultivar Embrapa 48, que foram esterilizadas seguindo o mesmo procedimento descrito nos experimentos anteriores. As sementes foram tratadas com os miPEPs na concentração de 10 µM, diluídos em água ultrapura. No tratamento controle, foram adicionados 50 µL de água Milli-Q autoclavada em cada poço. Para cada tratamento (controle, miPEP1 e miPEP2) foi utilizada uma placa para cultivo de células de 12 poços, com duas sementes por poço. Dessa forma, para cada tratamento foram feitas quatro réplicas com um *pool* de seis sementes por réplica. Após 3 horas de tratamento, foram adicionados 600 µL de água Milli-Q autoclavada em todos os poços. Foram utilizadas as mesmas condições de germinação descritas no tratamento com miR168 maduro. Em 30 HAE, os eixos embrionários foram removidos para análise fenotípica, como descrito no tópico 3.1.3.2. O experimento foi repetido para a extração de RNA total e análise da expressão gênica por RT-qPCR, seguindo o mesmo passo a passo feito anteriormente.

3.1.3.7. Aplicação e avaliação da expressão do miR168 em plântulas de soja

Para a avaliação da expressão do miR168 em plântulas de Embrapa 48 e BR-16 tratadas com 5 µg do miR168 maduro, a germinação foi realizada como detalhado no tópico 3.1.3.5. até o ponto de 3 HAE, e utilizando fotoperíodo de 12 horas. Para cada tratamento (controle e miR168 maduro) foram utilizadas 12 sementes, em ambas as cultivares. Em 3 HAE, as sementes foram transferidas para tubos de ensaio (150 × 25 mm) contendo 20 mL de meio de cultura Murashige e Skoog (MS) ½ força, suplementado com 30 g/L de sacarose e 7 g/L de ágar (Sigma-Aldrich). O pH foi ajustado para 5,8 e o meio de cultura foi esterilizado em autoclave por 15 minutos. Os tubos foram mantidos em incubadora B.O.D a 28°C e com fotoperíodo de 12/12 h por sete dias. Ao final dos sete dias de crescimento, as plântulas tiveram suas raízes e parte aérea seccionadas para análise fenotípica (comprimento (cm), massa fresca e

massa seca). Duas plântulas de cada tratamento foram utilizadas para extração de RNA total e análise da expressão gênica por RT-qPCR, como detalhado no tópico 3.1.3.3. O gene *Eukaryotic Translation Elongation Factor-2* (EEF2) foi utilizado como controle interno.

3.1.4. RESULTADOS

3.1.4.1. Estresse osmótico e salino afetam a germinação de soja

Para avaliar os efeitos do estresse osmótico e salino na germinação de soja, sementes das cultivares BR-16 e Embrapa 48 foram germinadas em diferentes concentrações de manitol e NaCl. Avaliamos as mudanças fenotípicas provocadas pelos tratamentos em relação à taxa de germinação (Figura 1), massa fresca (Figura 2), massa seca (Figura 3) e área do eixo embrionário (Figura 4). Todas as três concentrações de manitol aplicadas à cultivar BR-16 impactaram na taxa de germinação, principalmente a concentração de 300 mM, em que somente 20% das sementes germinaram (Figura 1A). Na cultivar Embrapa 48, o tratamento com manitol, de forma geral, atrasou a germinação. Mas, no ponto de 30 HAE, a porcentagem de sementes germinadas no controle e na maior concentração de manitol (300 mM) foi igual (aproximadamente 86%) (Figura 1B). O tratamento com NaCl, realizado com as mesmas concentrações utilizadas no tratamento com manitol, apresentou um impacto ainda maior na taxa de sementes germinadas, principalmente nas duas maiores concentrações. As sementes da cultivar BR-16 tratadas com 200 mM de NaCl apresentaram, aproximadamente, apenas 6% de sementes germinadas no ponto de 30 HAE. Já na concentração de 300 mM, nenhuma semente germinou (Figura 1C). A cultivar Embrapa 48 teve um desempenho um pouco melhor em relação à BR-16, mas também sofreu influência do estresse. Na concentração de 200 mM, a taxa de germinação foi de cerca de 40% ao fim da germinação, e na concentração de 300 mM, nenhuma semente foi capaz de germinar (Figura 1D).

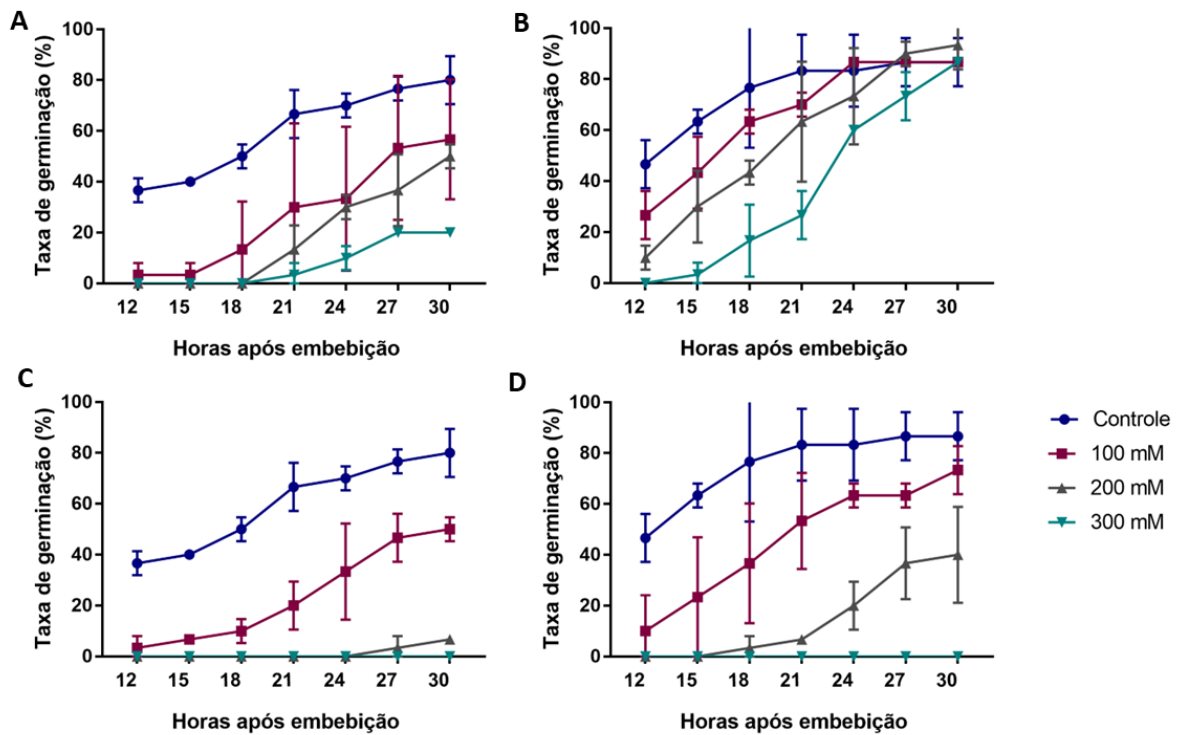


Figura 1: Taxa de germinação de sementes das cultivares BR-16 e Embrapa 48 não-tratadas e tratadas com manitol e NaCl nas concentrações de 100, 200 e 300 mM, em diferentes horas após a embebição. (A) Sementes da cultivar BR-16 tratadas com manitol. (B) Sementes da cultivar Embrapa 48 tratadas com manitol. (C) Sementes da cultivar BR-16 tratadas com NaCl. (D) Sementes da cultivar Embrapa 48 tratadas com NaCl.

Os tratamentos aplicados também afetaram a massa fresca dos eixos embrionários nas duas cultivares. No tratamento com manitol, a cultivar BR-16 apresentou a massa fresca dos eixos significativamente menor na concentração de 300 mM, em relação ao controle (Figura 2A). E na cultivar Embrapa 48, a massa fresca dos eixos nas concentrações de 200 mM e 300 mM também foi reduzida comparada ao controle (Figura 2A). No estresse salino, os eixos apresentaram massa fresca significativamente menor em relação ao controle em todas as concentrações utilizadas, em ambas as cultivares (Figura 2B).

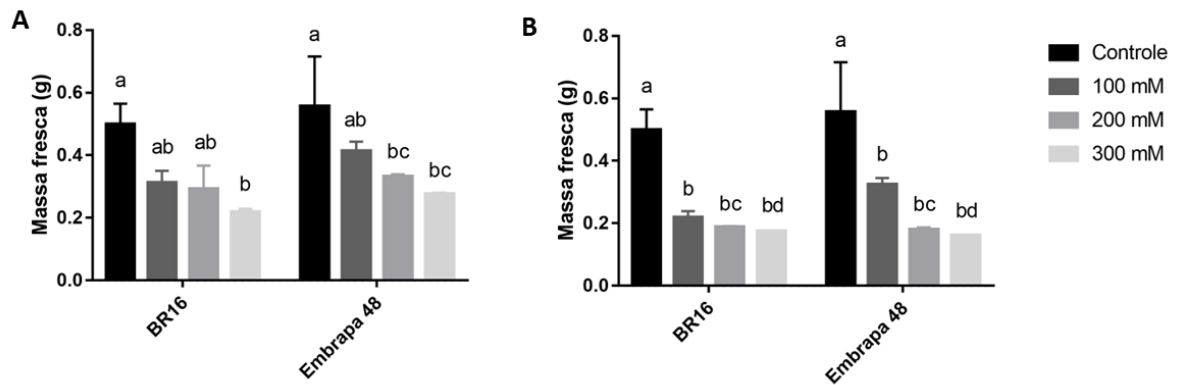


Figura 2: Massa fresca dos eixos embrionários no controle e nos tratamentos com manitol e NaCl, nas concentrações de 100, 200 e 300 mM, no ponto de 30 HAE. (A) Tratamento com manitol em ambas as cultivares. (B) Tratamento com NaCl em ambas as cultivares. As letras a, b, c e d indicam diferença significativa na massa fresca dos eixos embrionários controle e tratados, conforme teste de Tukey.

Em relação à massa seca dos eixos embrionários, não houve diferença estatística significativa entre o tratamento controle e o tratamento com manitol nas duas cultivares (Figura 3A). No entanto, o tratamento com NaCl causou redução significativa da massa seca dos eixos em BR-16, considerando todas as concentrações utilizadas, e em Embrapa 48, nas concentrações de 200 mM e 300 mM (Figura 3B).

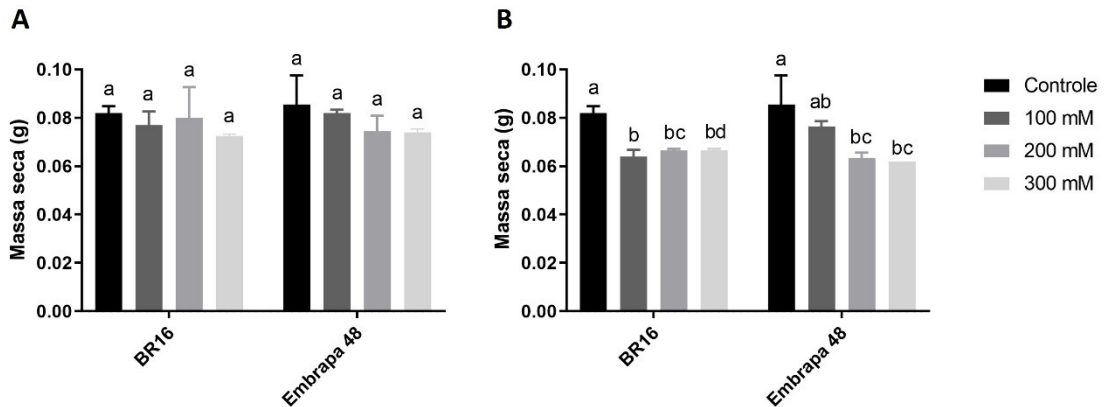


Figura 3: Massa seca dos eixos embrionários no controle e nos tratamentos com manitol e NaCl, nas concentrações de 100, 200 e 300 mM, no ponto de 30 HAE. (A) Tratamento com manitol em ambas as cultivares. (B) Tratamento com NaCl em ambas as cultivares. As letras a, b, c e d indicam diferença significativa na massa seca dos eixos embrionários controle e tratados, conforme teste de Tukey.

O estresse osmótico provocado pelo tratamento com 300 mM de manitol também causou a redução da área dos eixos embrionários em BR-16 (Figura 4A). Porém, em Embrapa 48 não houve mudança significativa entre as amostras tratadas e não tratadas (Figura 4B). Já o estresse salino afetou significativamente a área dos eixos de BR-16, nos tratamentos com 200 mM e 300 mM de NaCl (Figura 4C), bem como a área dos eixos de Embrapa 48, em todas as concentrações utilizadas (Figura 4D).

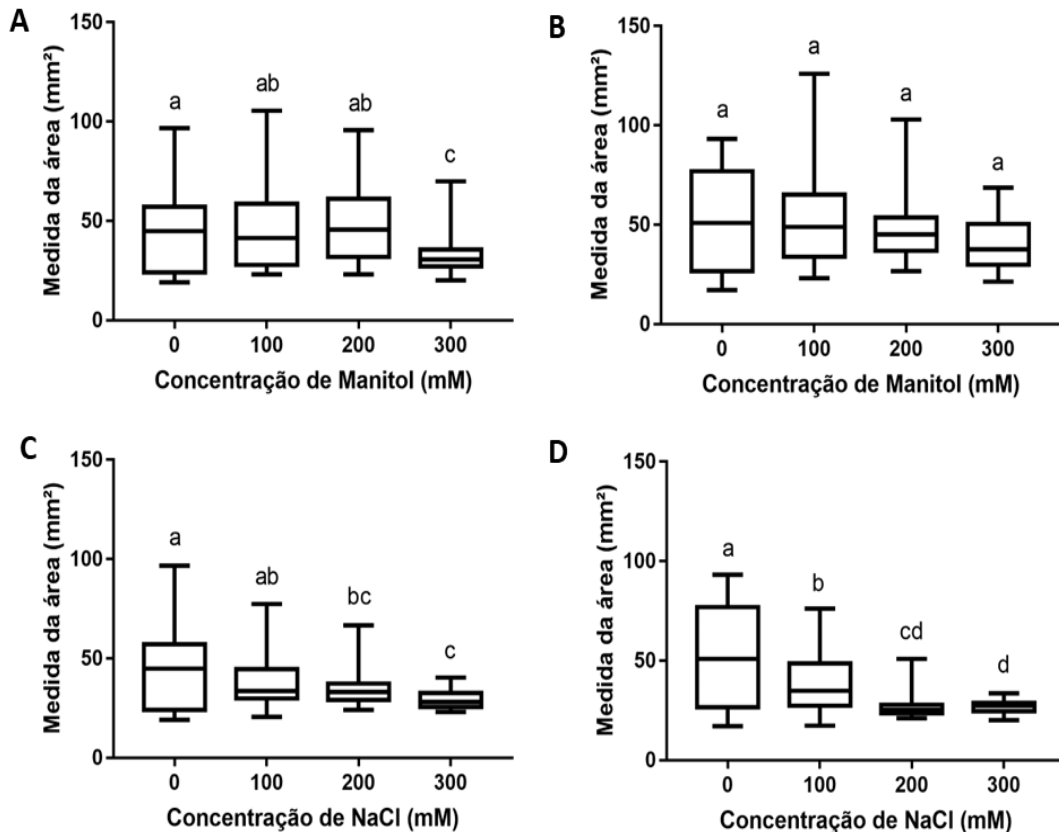
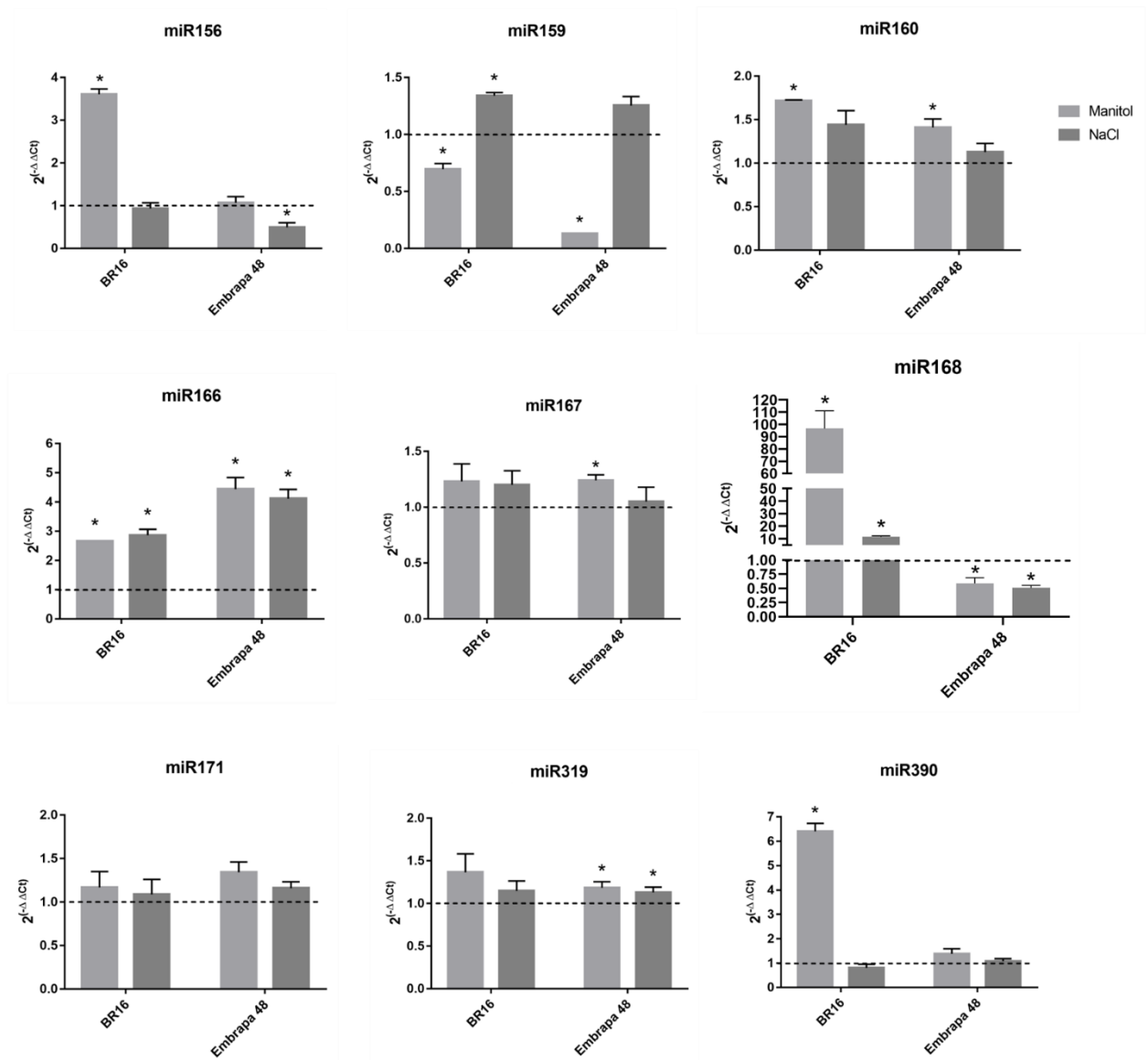


Figura 4: Medida da área dos eixos embrionários no controle e nos tratamentos com NaCl e manitol, nas concentrações de 100, 200 e 300 mM, no ponto de 30 HAE. (A) Tratamento com manitol em BR-16. (B) Tratamento com manitol em Embrapa 48. (C) Tratamento com NaCl em BR-16. (D) Tratamento com NaCl em Embrapa 48. As letras a, b, c e d indicam diferença significativa na área dos eixos embrionários controle e tratados, conforme teste de Tukey.

3.1.4.2. Estresse osmótico e salino podem alterar os perfis de expressão de miRNAs durante a germinação de soja

Após as análises descritas anteriormente, foram selecionadas as concentrações de 300 mM de manitol e 100 mM de NaCl para os experimentos seguintes, que incluíram a análise de expressão de miRNAs por RT-qPCR. Foram observados miRNAs com diferença significativa de expressão entre os tratamentos e o controle (Figura 5). Com exceção do miR171, todos os miRNAs apresentaram mudança de expressão em resposta a pelo menos um dos tratamentos. A maior parte dos miRNAs exibiu expressão aumentada em um ou ambos os estresses, como: miR160, miR166, miR167, miR319, miR390, miR394, miR398, miR408, miR482, miR1507 e miR3522. Desses, o miR166, miR394 e miR398 tiveram a expressão aumentada nos dois estresses, em ambas as cultivares. Curiosamente, alguns miRNAs apresentaram perfis de expressão contrastantes entre as duas cultivares,

como o miR156, que foi *up*-regulado no estresse osmótico em BR-16, e em Embrapa 48 não houve mudança significativa da expressão no mesmo estresse. O miR168 chamou uma atenção ainda maior, por ter sido o miRNA com maior nível de expressão entre todos os analisados, e por ter tido uma alta expressão sob estresse osmótico em BR16, enquanto foi *down*-regulado em Embrapa 48 na mesma condição.



(continua)

(continuação)

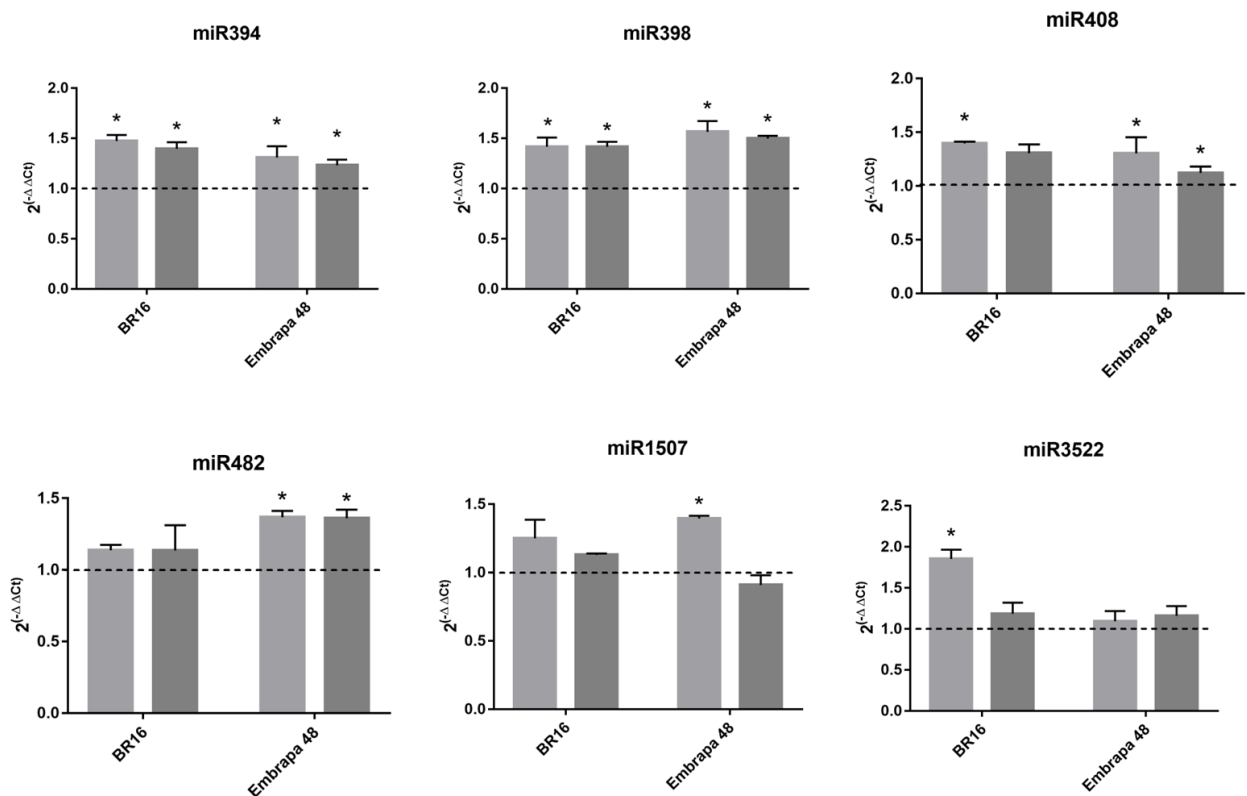


Figura 5: Perfis de expressão de 15 miRNAs relacionados a estresses durante a germinação de sementes de soja, tratadas com manitol a 300 mM e NaCl a 100 mM. A linha tracejada representa a expressão dos miRNAs nas sementes controle. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$).

3.1.4.3. Efeitos do tratamento com miR168 exógeno

Considerando a grande discrepância nos perfis de expressão do miR168 entre as duas cultivares, em condição de estresse osmótico, buscamos investigar a possibilidade de aplicação exógena do miR168 maduro em sementes da cultivar Embrapa 48, em que foi *down*-regulado, e avaliar os efeitos desse tratamento. Inicialmente, foi feita a aplicação de 1 μ g e 5 μ g do miR168 maduro nas sementes em 0 HAE. Em 30 HAE, os eixos embrionários das sementes tratadas tanto com 1 μ g, quanto com 5 μ g do miR168 maduro apresentaram a expressão do miR168 significativamente maior em relação ao controle (Figura 6).

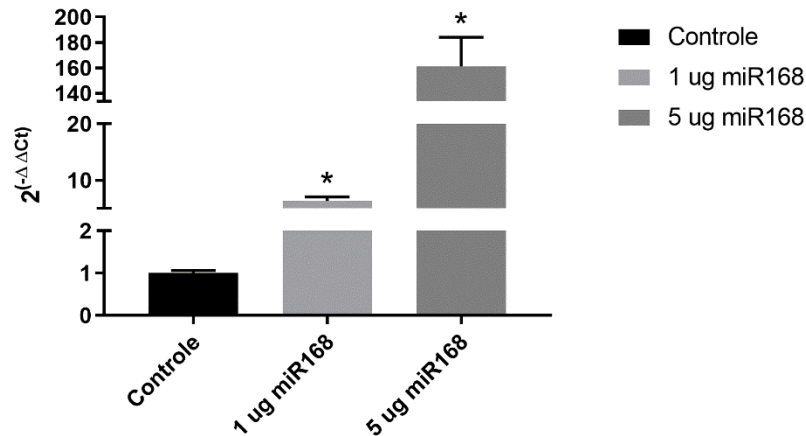


Figura 6: Perfil de expressão do miR168 em eixos embrionários da cultivar Embrapa 48, em 30 HAE, após aplicação de 1 µg e 5 µg do miR168 maduro. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$).

Em seguida, foi feita uma análise fenotípica para avaliar a massa fresca, massa seca e área dos eixos embrionários nos seguintes tratamentos: controle, manitol a 300 mM, 5 µg do miR168, e 5 µg do miR168 + manitol a 300 mM. Na massa fresca, não houve diferença estatística significativa entre as amostras controle e as tratadas com o miR168 maduro. No entanto, a massa fresca dos eixos tratados com miR168 foi significativamente maior em relação aos tratados com manitol e miR168+manitol. Também não houve diferença entre as amostras tratadas com miR168+manitol e as tratadas somente com manitol (Figura 7A). O mesmo foi observado na análise de massa seca (Figura 7B) e área dos eixos (Figura 7C).

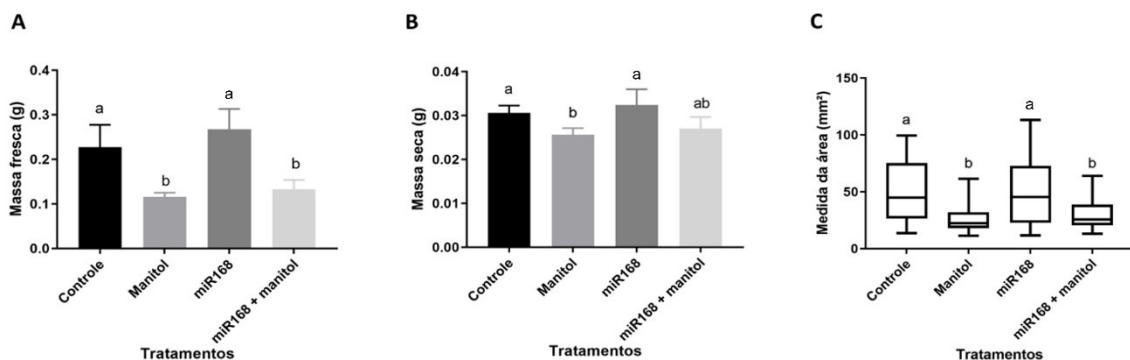


Figura 7: Análises fenotípicas após tratamentos com manitol a 300 mM, 5 µg do miR168, e 5 µg do miR168 + manitol a 300 mM em Embrapa 48. (A) Massa fresca dos eixos em 30 HAE. (B) Massa seca dos eixos em 30 HAE. (C) Área dos eixos em 30 HAE. As letras a, b, c e d indicam diferença significativa entre os tratamentos, conforme teste de Tukey.

Por meio de RT-qPCR, buscamos avaliar a expressão do miR168 e de dois dos seus genes-alvo AGO1 preditos (Glyma.16G217300 e Glyma.09G167100) (Tabela S2), mediante o tratamento conjunto com o miR168 exógeno e manitol. O miR168 exibiu expressão aumentada no tratamento com 5 μg do miR168 exógeno, com e sem manitol, reforçando o que foi mostrado na Figura 6. No entanto, quando o manitol foi administrado, a expressão do miR168 foi menor em comparação com o tratamento exclusivo com miR168 (Figura 8A). Os dois genes AGO1 demonstraram perfis de expressão semelhantes, com aumento de expressão nos tratamentos contendo 300 mM de manitol, 5 μg de miR168 e na combinação de miR168 e manitol, tendo exibido a maior expressão nesse último tratamento (Figura 8B-C).

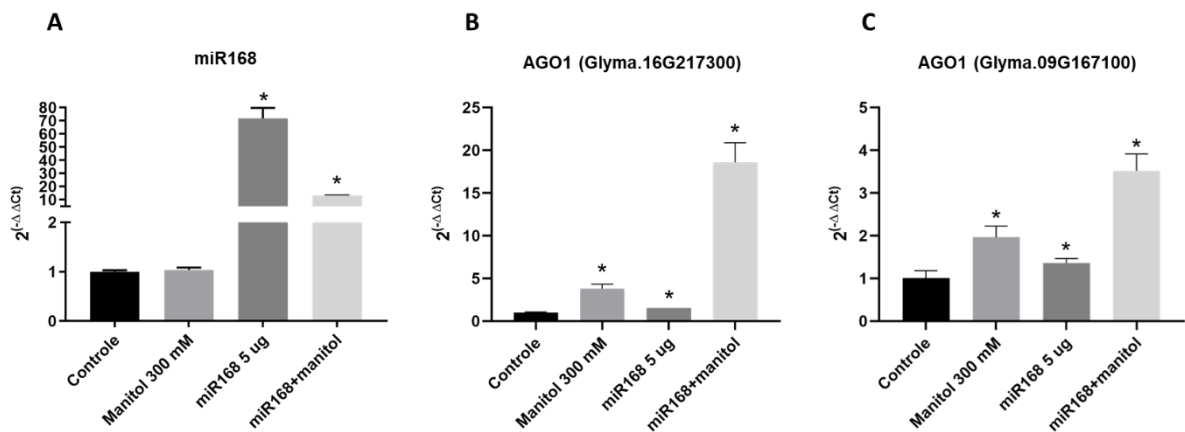


Figura 8: Perfil de expressão do miR168 e seus alvos AGO1 (Glyma.16G217300 e Glyma.09G167100), em eixos embrionários da cultivar Embrapa 48, em 30 HAE, após tratamento com 5 μg do miR168 maduro e manitol a 300 mM. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$).

3.1.4.4. Tratamento com miPEPs sintéticos associados ao miR168

Além da aplicação do miR168 maduro, foi realizada a predição e aplicação de potenciais miPEPs derivados do miR168 primário. Buscamos analisar se o tratamento com os dois miPEPs preditos (MRLRWGF - miPEP1 e MPCDCVWGF - miPEP2) poderia aumentar a expressão do miR168 e causar mudanças fenotípicas em Embrapa 48. Os resultados das análises de massa fresca, massa seca e área dos eixos em 30 HAE não apresentaram diferença significativa entre o tratamento com os dois miPEPs e o controle (Figura 9). Por meio de RT-qPCR, foram analisados os perfis de expressão do miR168 e dos seus genes-alvo AGO1, após a aplicação dos miPEPs.

A expressão do miR168 não foi aumentada pelos miPEPs, assim como a expressão dos genes-alvo não apresentou alteração (Figura 10).

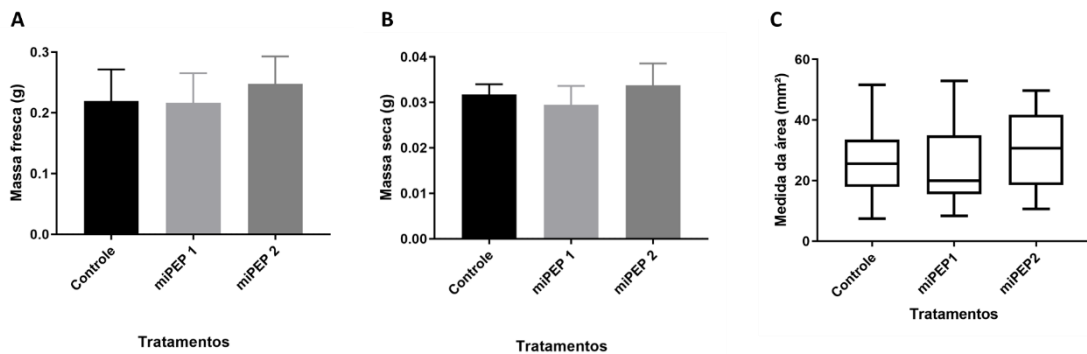


Figura 9: Análises fenotípicas após tratamento com miPEPs em Embrapa 48. (A) Massa fresca dos eixos em 30 HAE. (B) Massa seca dos eixos em 30 HAE. (C) Área dos eixos em 30 HAE.

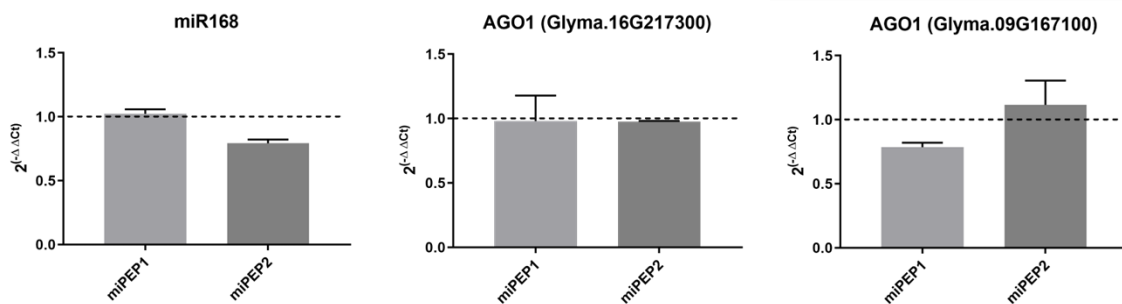


Figura 10: Perfis de expressão do miR168 e seus alvos AGO1 (Glyma.16G217300 e Glyma.09G167100), após tratamento com miPEPs em Embrapa 48. A linha tracejada representa a expressão dos genes nas amostras controle.

3.1.4.5. Aplicação do miR168 em plântulas de soja

Por fim, buscamos investigar se o miR168 exógeno aplicado em 0 HAE poderia estar ativo em plântulas de sete dias, e se causaria modificações fenotípicas em BR-16 e Embrapa 48. Após analisar o comprimento, massa fresca e massa seca das plântulas, foi observado que o tratamento com o miR168 maduro não alterou as características fenotípicas das plântulas de ambas as cultivares, não havendo diferenças significativas entre as plântulas tratadas e não-tratadas (Figura 11).

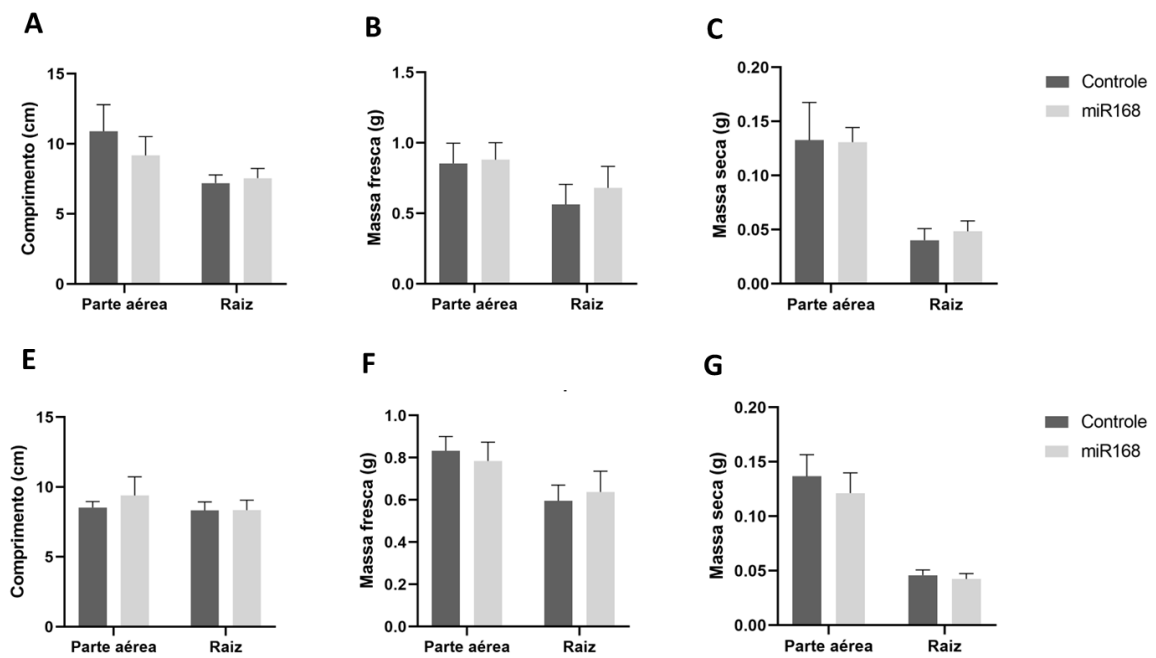


Figura 11: Análise fenotípica de plântulas de sete dias das cultivares BR16 e Embrapa 48 após tratamento com miR168 maduro. (A) Comprimento, (B) Massa fresca e (C) Massa seca de plântulas controle e tratadas da cultivar BR16. (E) Comprimento, (F) Massa fresca e (G) Massa seca de plântulas controle e tratadas da cultivar Embrapa 48.

No entanto, por meio de RT-qPCR, verificamos que a expressão do miR168 foi significativamente aumentada em raízes de plântulas tratadas com o miR168 exógeno, em ambas as cultivares (Figura 12). Além disso, os dois genes-alvo AGO1 (Glyma.16G217300 e Glyma.09G167100) exibiram expressão aumentada em BR-16 (Figura 12A). Em Embrapa 48, a expressão dos dois genes também exibiu diferença significativa em relação ao controle, sendo o gene Glyma.16G217300 *up*-regulado e o gene Glyma.09G167100 *down*-regulado (Figura 12B).

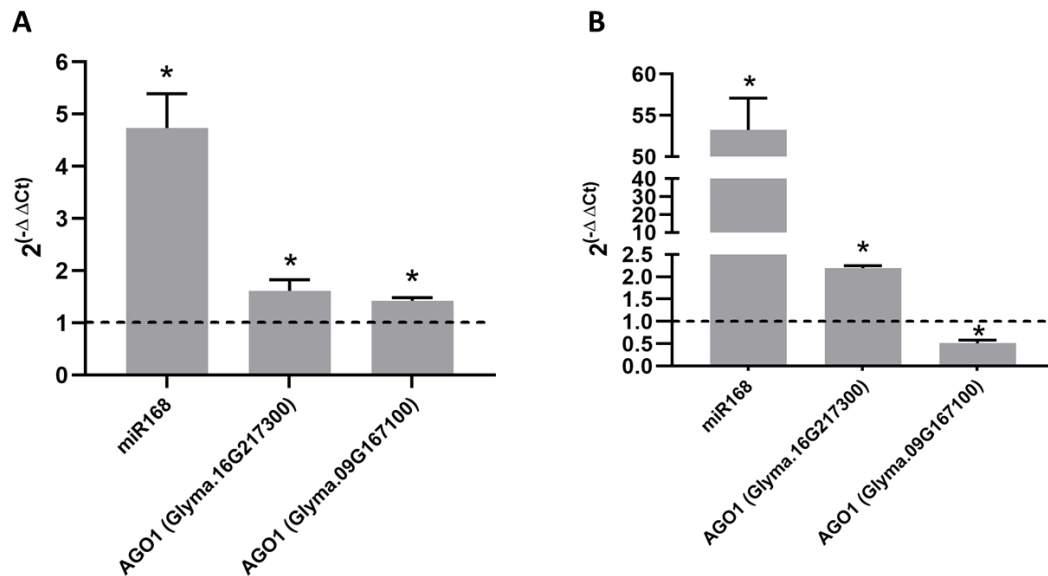


Figura 12: Perfil de expressão do miR168 e dos genes-alvo AGO1 (Glyma.16G217300 e Glyma.09G167100) em raízes de BR16 (A) e Embrapa 48 (B) tratadas com miR168 exógeno. Os asteriscos (*) indicam os valores de expressão significativamente diferentes entre o controle e os tratamentos, conforme teste t ($P < 0,05$).

3.1.5. DISCUSSÃO

O estresse salino é um dos estresses abióticos mais comuns, e impacta a produtividade e qualidade de diversas plantas cultivadas. O processo germinativo é particularmente uma das fases do desenvolvimento vegetal mais vulneráveis ao estresse salino, pois prejudica a absorção de água durante embebição e turgescência das sementes (DE LA REGUERA et al., 2020; TARCHOUN et al., 2022). Igualmente relevante, a seca é um dos estresses ambientais mais severos, que pode ocorrer devido à dinâmica de temperatura, intensidade da luz e baixa pluviosidade. A seca afeta a germinação, e conseqüentemente, o crescimento da planta e rendimento de culturas (SELEIMAN et al., 2021; THABET et al., 2018). Um número crescente de miRNAs vêm sendo relatados como responsivos a estresses ambientais, e auxiliam no balanço entre o desenvolvimento das plantas e a resposta a estresses abióticos (ZHANG et al., 2022). No presente trabalho, buscamos analisar os efeitos desses estresses na germinação de soja e na regulação de miRNAs.

O nosso estudo se iniciou com a avaliação das características fenotípicas de sementes germinadas com diferentes concentrações de NaCl e manitol, o qual é comumente utilizado para induzir a seca por meio do estresse osmótico

(NIKONOROVA et al., 2018; SAADAOUI et al., 2023). Foram utilizadas duas cultivares de soja com tolerâncias diferentes à seca: BR-16 (sensível) e Embrapa 48 (tolerante). Como esperado, ambos os estresses, de forma geral, causaram um impacto maior na cultivar sensível à seca, o que ficou mais evidente no tratamento com manitol. As sementes de BR-16 tratadas com a menor concentração de manitol (100 mM) apresentaram uma taxa de germinação de 56%, enquanto 80% das sementes não-tratadas germinaram (Figura 1A), o que demonstra que mesmo em níveis menores o manitol é capaz de induzir estresse osmótico nessa cultivar. No estresse salino, a cultivar sensível também apresentou um impacto maior na taxa de germinação quando comparada à cultivar tolerante (Figura 1C-D). No entanto, quando se considera a massa fresca dos eixos embrionários, observa-se que o tratamento com NaCl afetou igualmente as duas cultivares (Figura 2B). De forma semelhante, na medida da área dos eixos embrionários, a cultivar Embrapa 48 apresentou eixos com área reduzida em todas as concentrações de NaCl (Figura 4D). Isso indica que, em condições de estresse salino, mesmo a cultivar tolerante à seca sofre impactos nas características fenotípicas. Um estudo feito com outras quatro cultivares de soja brasileiras (Pioneira, Conquista, Pintado e Xingu) demonstrou que os tratamentos com NaCl e manitol afetam tanto a germinação, quanto o desenvolvimento de plântulas (NETO et al., 2004). Também foi visto que o uso de NaCl causou um impacto maior na porcentagem de sementes germinadas em comparação com o manitol (NETO et al., 2004).

Após os tratamentos com NaCl e manitol durante a germinação, investigamos se os perfis de expressão de 15 miRNAs responsivos a estresses abióticos em plantas foram modificados. Dos 15 miRNAs analisados, 14 apresentaram mudanças nos padrões de expressão em pelo menos uma das cultivares (Figura 5). Alguns desses miRNAs exibiram perfis de expressão contrastantes entre as duas cultivares, como: o miR156, que apresentou expressão aumentada no estresse osmótico em BR-16, não tendo alteração em Embrapa 48. E ao mesmo tempo foi *down*-regulado no estresse salino em Embrapa 48, e em BR-16 não houve mudança de expressão em relação ao controle. Já os miR390 e miR3522 foram *up*-regulados somente no tratamento com manitol em BR-16, não havendo mudanças de expressão nos outros tratamentos. E o miR168 apresentou aumento de expressão em BR-16 em ambos os estresses, enquanto em Embrapa 48 ocorreu o contrário, sendo *down*-regulado nos dois tratamentos. Perfis de expressão contrastantes entre as duas cultivares também

foram demonstrados por Kulcheski e colaboradores (2011). Nesse estudo, foram avaliados os padrões de expressão de miRNAs em condição de déficit hídrico, considerando tecidos de raiz de BR-16 e Embrapa 48. Foi visto que a maioria dos miRNAs exibiu aumento de expressão na cultivar sensível, e foram *down*-regulados na cultivar tolerante. Isso indica que diferentes genótipos de uma mesma espécie podem apresentar perfis de expressão de miRNAs distintos em resposta a estresses (KULCHESKI et al., 2011).

Dentre os miRNAs que exibiram perfis de expressão contrastantes entre as cultivares, o miR168 se destacou por ter tido a maior expressão entre todos os miRNAs. Observou-se uma expressão significativamente aumentada em BR-16 após o tratamento com manitol, enquanto na cultivar tolerante a expressão foi reduzida. Em *Arabidopsis*, já foi demonstrado que o miR168 responde tanto à seca, a partir do tratamento com manitol, como à salinidade e ao frio (LIU et al., 2008). O miR168 tem um importante papel na regulação da AGO1, componente central do complexo RISC, que direciona o silenciamento dos genes-alvo. Dessa forma, a homeostase da AGO1 é mantida por meio de um ciclo de feedback, envolvendo a coexpressão do miR168 e AGO1, e a interação de ambos no nível pós-transcricional (LI et al., 2012; VAUCHERET; MALLORY; BARTEL, 2006). Em nosso estudo, buscamos analisar se a aplicação exógena do miR168 em Embrapa 48, em que foi *down*-regulado, seria capaz de aumentar sua expressão e se modularia a expressão de genes-alvo AGO1. Foi feita a predição dos alvos pelo psRNATarget (Tabela S2), e foram selecionados dois dos genes AGO1 para análise (Glyma.16G217300 e Glyma.09G167100). Já foi demonstrado que a aplicação de miRNAs exógenos em plantas é capaz de regular a expressão de genes-alvo de interesse, como descrito por Betti et al. (2021). Nesse estudo, os miRNAs foram sintetizados seguindo as características que os miRNAs endógenos possuem, isto é, com uracila em sua composição e com um grupo metil na extremidade 3' (BETTI et al., 2021). Em nosso trabalho buscamos investigar se o miRNA contendo timina em sua composição e sem adição do grupo metil na extremidade 3' poderia ser utilizado como uma alternativa. Apesar de o tratamento realizado com o miR168 exógeno não ter causado modificações fenotípicas significativas, foi capaz de aumentar a expressão do miR168 tanto em eixos embrionários em 30 HAE (Figura 6), como em raízes de plântulas com sete dias (Figura 12). Além disso, nas amostras de eixos embrionários e raízes de plântulas tratadas exogenamente com o miR168, os dois genes-alvo AGO1 analisados

apresentaram perfis de expressão significativamente diferentes das amostras controle, indicando que o miR168 aplicado foi capaz de regular a expressão desses genes (Figura 8 e Figura 12). Nas amostras de eixos embrionários, os genes AGO1 foram *up*-regulados em todos os tratamentos (300 mM de manitol, 5 µg de miR168 e na combinação de miR168 e manitol) (Figura 8). E em raízes de plântulas com sete dias, ambos os genes AGO1 foram *up*-regulados em BR-16, e em Embrapa 48, o gene Glyma.16G217300 apresentou aumento de expressão, enquanto o Glyma.09G167100 foi *down*-regulado (Figura 12). Esses dados indicam que a atividade do miR168 aplicado antes da germinação permaneceu em plântulas após sete dias. Os resultados obtidos foram semelhantes à regulação de AGO1 observada em *Arabidopsis* em situação de seca. Um estudo feito por Li e colaboradores (2012) mostrou que o miR168 controla a homeostase de AGO1 em resposta a estresses abióticos, incluindo o tratamento com desidratação. Foi visto que o miR168 foi induzido em condição de seca, assim como a atividade de transcrição da AGO1 foi aumentada sob o mesmo tratamento (Li et al., 2012), sendo semelhante aos perfis de expressão de AGO1 apresentados nas Figuras 8 e 12. Estudos mais aprofundados são necessários para validar a eficácia da aplicação do miRNA maduro com as características aqui consideradas, e as possíveis aplicações biotecnológicas.

Além da aplicação exógena do miR168, também realizamos o tratamento com potenciais miPEPs derivados do pri-miR168. Dois miPEPs preditos (MRLRWGF - miPEP1 e MPCDCVWGF - miPEP2) foram sintetizados e aplicados em sementes de Embrapa 48 em 0 HAE. Diferente do que foi esperado, as análises fenotípicas realizadas em 30 HAE revelaram que o tratamento não alterou a massa fresca e seca, bem como a área dos eixos embrionários (Figura 9). Os miPEPs também não aumentaram a expressão do miR168 maduro, nem dos seus alvos (Figura 10). Trabalhos anteriores demonstraram que, em vários pri-miRNAs, a primeira ORF após a região do sítio de início da transcrição (TSS) foi capaz de produzir um miPEP funcional, que aumentou a expressão do seu miRNA associado (LAURESSERGUES et al., 2022; CHEN et al., 2020). Em nosso trabalho, as ORFs selecionadas para a predição dos miPEPs também seguiram o critério de localização após a região TSS, porém, não foi observado o efeito esperado. Esses dados sugerem que a codificação de um miPEP a partir da primeira ORF após a TSS não parece estar atrelada a sua funcionalidade no aumento da expressão do miRNA associado como proposto por Laressergues et al., 2022.

3.1.6. CONCLUSÕES

Com base nos resultados apresentados, podemos concluir que:

- A germinação de soja é afetada pelos estresses salino e osmótico, sendo que o tratamento com NaCl provocou maior impacto, tanto na taxa de germinação, como na massa fresca, massa seca e área dos eixos embrionários em ambas as cultivares;
- A expressão de miRNAs é modulada por estresse salino e osmótico durante a germinação de soja;
- Os miRNAs possuem perfis de expressão contrastantes entre a cultivar suscetível e a tolerante;
- O tratamento com o miR168 exógeno é capaz de aumentar sua expressão na cultivar em que foi *down*-regulado;
- A expressão do miR168 exógeno permaneceu aumentada em plântulas após sete dias de tratamento;
- A aplicação do miR168 em 0 HAE foi capaz de regular a expressão de dois genes-alvo AGO1 em eixos embrionários após 30 horas de embebição e em plântulas com sete dias;
- O tratamento com dois potenciais miPEPs relacionados ao pri-miR168 não causou mudanças fenotípicas significativas na germinação de soja, e não alterou a expressão do miR168 maduro;
- A produção de miPEPs a partir da primeira ORF após a região TSS não parece ser uma regra geral para todas as ORFs.

3.1.7. REFERÊNCIAS BIBLIOGRÁFICAS

ABRÀMOFF, M. D.; MAGALHÃES, P. J.; RAM, S. J. Image processing with imageJ. **Biophotonics International**, v. 11, n. 7, p. 36–41, 2004.

ARAÚJO, P. M. DE; GRATIVOL, C. In silico identification of candidate miRNA-encoded Peptides in four Fabaceae species. **Computational Biology and Chemistry**, v. 97, 2022.

BARRERA-FIGUEROA, B. E. et al. Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. **BMC Plant Biology**, v. 11, 2011.

BETTI, F. et al. Exogenous miRNAs induce post-transcriptional gene silencing in plants. **Nature Plants**, v. 7, n. 10, p. 1379–1388, 2021.

CHEN, Q. J. et al. A mirna-encoded small peptide, vvi-miPEP171d1, regulates adventitious root formation. **Plant Physiology**, v. 183, n. 2, p. 656–670, 2020.

DAI, X.; ZHUANG, Z.; ZHAO, P. X. PsRNATarget: A plant small RNA target analysis server (2017 release). **Nucleic Acids Research**, v. 46, n. W1, p. W49–W54, 2018.

DAS, S. S. et al. Small RNA mediated regulation of seed germination. **Frontiers in Plant Science**, v. 6, 2015.

DE LA REGUERA, E. et al. The effects of saltwater intrusion on germination success of standard and alternative crops. **Environmental and Experimental Botany**, v. 180, 2020.

JUNG, H. J.; KANG, H. Expression and functional analyses of microRNA417 in *Arabidopsis thaliana* under stress conditions. **Plant Physiology and Biochemistry**, v. 45, n. 10–11, p. 805–811, 2007.

KIM, J. Y. et al. MicroRNA402 affects seed germination of *Arabidopsis thaliana* under stress conditions via targeting DEMETER-LIKE Protein3 mRNA. **Plant and Cell Physiology**, v. 51, n. 6, p. 1079–1083, 2010a.

KIM, J. Y. et al. Overexpression of microRNA395c or 395e affects differently the seed germination of *Arabidopsis thaliana* under stress conditions. **Planta**, v. 232, n. 6, p. 1447–1454, 2010b.

KULCHESKI, F. R. et al. Identification of novel soybean microRNAs involved in abiotic

and biotic stresses. **BMC Genomics**, v. 12, 2011.

LAURESSERGUES, D. et al. Characterization of plant microRNA-encoded peptides (miPEPs) reveals molecular mechanisms from the translation to activity and specificity. **Cell Reports**, v. 38, n. 6, 2022.

LI, H. et al. Characterization of the stress associated microRNAs in Glycine max by deep sequencing. **BMC Plant Biology**, v. 11, 2011.

LI, W. et al. Transcriptional regulation of arabidopsis MIR168a and ARGONAUTE1 homeostasis in abscisic acid and abiotic stress responses. **Plant Physiology**, v. 158, n. 3, p. 1279–1292, 2012.

LIU, H. H. et al. Microarray-based analysis of stress-regulated microRNAs in Arabidopsis thaliana. **Rna**, v. 14, n. 5, p. 836–843, 2008.

LIU, P. P. et al. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. **Plant Journal**, v. 52, n. 1, p. 133–146, 2007.

LUJÁN-SOTO, E.; DINKOVA, T. D. Time to wake up: Epigenetic and small-RNA-mediated regulation during seed germination. **Plants**, v. 10, n. 2, p. 1–19, 2021.

NETO, N. B. M. et al. Water stress induced by mannitol and sodium chloride in soybean cultivars. **Brazilian Archives of Biology and Technology**, v. 47, n. 4, p. 521–529, 2004.

NIKONOROVA, N. et al. Early mannitol-triggered changes in the Arabidopsis leaf (phospho)proteome reveal growth regulators. **Journal of Experimental Botany**, v. 69, n. 19, p. 4591–4607, 2018.

RAO, X. et al. An improvement of the $2^{-\Delta\Delta CT}$ method for quantitative real-time polymerase chain reaction data analysis. **Biostatistics, bioinformatics and biomathematics**, v. 3, n. 3, p. 71–85, 2013.

REYES, J. L.; CHUA, N. H. ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. **Plant Journal**, v. 49, n. 4, p. 592–606, 2007.

SAADAoui, W. et al. Effects of drought stress induced by D-Mannitol on the germination and early seedling growth traits, physiological parameters and

phytochemicals content of Tunisian squash (*Cucurbita maxima*Duch.) landraces. **Frontiers in Plant Science**, v. 14, 2023.

SELEIMAN, M. F. et al. Drought stress impacts on plants and different approaches to alleviate its adverse effects. **Plants**, v. 10, n. 2, p. 1–25, 2021.

TARCHOUN, N. et al. The Effects of Salt Stress on Germination, Seedling Growth and Biochemical Responses of Tunisian Squash (*Cucurbita maxima* Duchesne) Germplasm. **Plants**, v. 11, n. 6, 2022.

THABET, S. G. et al. Genetic basis of drought tolerance during seed germination in barley. **PLoS ONE**, v. 13, n. 11, 2018.

VARKONYI-GASIC, E.; HELLENS, R. P. Quantitative Stem-Loop RT-PCR for Detection of MicroRNAs. In: KODAMA, H.; KOMAMINE, A. (Eds.). . **RNAi and Plant Gene Function Analysis**. [s.l.] Humana Press, 2011. p. 145–157.

VAUCHERET, H.; MALLORY, A. C.; BARTEL, D. P. AGO1 Homeostasis Entails Coexpression of MIR168 and AGO1 and Preferential Stabilization of miR168 by AGO1. **Molecular Cell**, v. 22, n. 1, p. 129–136, 2006.

ZHANG, Y. et al. Non-coding RNAs fine-tune the balance between plant growth and abiotic stress tolerance. **Frontiers in Plant Science**, v. 13, 2022.

ZHENG, Y. et al. Small RNA profiles in soybean primary root tips under water deficit. **BMC Systems Biology**, v. 10, 2016.

3.2. Capítulo 2: Identificação e evolução da família gênica CLASSY em genomas de plantas

CLASSY gene family finding and evolution in plants genomes

Paula Machado de Araújo^a, Arthur Gruber^b, Liliane Santana Oliveira^b, Sara Sangi^c, Geovanna Vitória Olimpio^a, Felipe Cruz Paula^a, and Clícia Grativol^{a,*}

^a Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brazil

^b Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brazil

^c Laboratório de Biologia Celular e Tecidual, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brazil

*Corresponding author: cgrativol@uenf.br; Tel.: +55 22 2739-7107

Full postal address: Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto Lamego, 2000, P5, 228, Parque Califórnia, Campos dos Goytacazes, Rio de Janeiro, Brazil.

3.2.1. ABSTRACT

RNA-directed DNA methylation (RdDM) is an epigenetic mechanism involved in several biological processes in plants, such as response to biotic and abiotic stresses, developmental regulation, and silencing of transposable elements. RdDM leads to *de novo* DNA methylation via small RNAs, which requires a complex machinery with multiple components, including the chromatin remodeling protein CLASSY (CLSY). The CLSY family controls both global and locus-specific regulation of DNA methylation and has been identified only in the model plant *Arabidopsis thaliana*. In this work, we sought to identify CLSY proteins in different plants, and analyzed their phylogeny in moss (bryophyte),

33 basal angiosperm, basal eudicot, monocots, and eudicots. We identified
34 candidate CLSYs in 11 of the 12 species investigated. Phylogenetic analysis
35 indicated that CLSY proteins are divided into two main groups, one group being
36 similar to *Arabidopsis* CLSY1 and CLSY2, while the other resembles *Arabidopsis*
37 CLSY3 and CLSY4. Gene duplication analysis demonstrated that whole-genome
38 duplication/segmental duplication events contributed to the expansion of the
39 CLSY family in soybean. Using RT-qPCR, we evaluated the expression of CLSY
40 and five other genes related to epigenetic regulation during soybean germination
41 subjected to salt and osmotic stress. It was observed that the expression of these
42 genes can be modulated by stress, and their expression profile varied between
43 different soybean cultivars. The data generated in this study enhance our
44 comprehension of the evolutionary dynamics within the CLSY family in plants.
45 Furthermore, it furnishes insights into the CLSY response to abiotic stress.

46 **Keywords:** Epigenetic regulation; RNA-directed DNA methylation; CLSY1-4;
47 Phylogeny.

48
49 **Abbreviations:** AGO, ARGONAUTE; CLSY, CLASSY; DCL3, DICER-LIKE 3;
50 DRD1, DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1; DRM2,
51 DOMAINS REARRANGED METHYLTRANSFERASE 2; dsRNA, double-
52 stranded RNA; HMM, Hidden Markov Models; Pol, RNA Polymerase; RdDM,
53 RNA-directed DNA methylation; RDR2, RNA-DEPENDENT RNA POLYMERASE
54 2; ROS1, REPRESSOR OF SILENCING 1; RT-qPCR, Reverse transcription-
55 quantitative polymerase chain reaction; SHH1, SAWADEE HOMEODOMAIN
56 HOMOLOG 1; siRNA, small interfering RNA; sRNA, small RNA; ssRNA, single-
57 stranded RNA; TE, transposable element; WGD, whole-genome duplication.

59 3.2.2. INTRODUCTION

60 Epigenetic marks are characterized by changes in genomes that do not
61 alter the primary DNA sequence, and that can be inherited through cell division
62 (Henderson and Jacobsen, 2007). Different epigenetic pathways in plants
63 contribute to phenotypic plasticity and survival under unpredictable
64 environmental conditions (Pikaard and Scheid, 2014). The mechanisms of DNA
65 methylation, histone modification, and small RNA-guided DNA methylation can

66 act individually or together to promote tolerance to biotic and abiotic stresses
67 (Grativol et al., 2012). Also, epigenetic regulation plays an important role in
68 various stages of plant development, such as seed development, germination,
69 fruit ripening, and sexual and asexual reproduction (Brukhin and Albertini, 2021;
70 Kumar and Mohapatra, 2021). Plants with aberrant methylation patterns may
71 exhibit developmental abnormalities (Zhang et al., 2018). In fact, epigenetic
72 modifications provide an additional level on genetic regulation, that impacts the
73 growth, environmental adaptation, and the evolutionary history of plants
74 (Hemenway and Gehring, 2023; Pikaard and Scheid, 2014).

75 RNA-directed DNA methylation (RdDM) is an epigenetic pathway, unique
76 to plants, in which small RNAs (sRNAs) guide *de novo* DNA methylation
77 (Erdmann and Picard, 2020). RdDM was first reported in transgenic tobacco
78 plants infected with viroids, consisting of a circular non-coding RNA
79 (Wassenegger et al., 1994). RdDM is involved in the response to biotic stresses,
80 such as those caused by bacteria, viruses, and fungi (Erdmann and Picard, 2020;
81 Matzke and Mosher, 2014), and abiotic stresses, such as salt (Xu et al., 2015)
82 and heat stress (Popova et al., 2013). Plants are subjected to programmed
83 chromatin changes in response to the environment, and the RdDM pathway is
84 crucial to this process. An example of the importance of RdDM pathway to plant
85 stress response is the MutS Homolog1 (MSH1) system. The disruption of the
86 MSH1 gene impacts the stress response in *Arabidopsis thaliana* by inducing
87 epigenomic reprogramming via RdDM (Kundariya et al., 2022).

88 Another important function of RdDM is the repression of transposable
89 element (TE) activity. Without TE silencing by the RdDM pathway, TEs can be
90 inserted into genes or promoters, which can affect gene expression or cause
91 mutations in proteins (Erdmann and Picard, 2020). Therefore, RdDM helps
92 maintain genomic stability (He et al., 2021), especially in plants with high TE
93 content, such as maize, where approximately 85% of the genome is composed
94 of TE (Schnable et al., 2009). Furthermore, the RdDM pathway contributes to the
95 regulation of stress-induced TEs activation. One example is the retrotransposon
96 *ONSEN*, which is upregulated during heat stress in *Arabidopsis*, but is repressed
97 by sRNAs associated with RdDM. *ONSEN* transposition was only observed in
98 stressed mutant plants with the RdDM pathway compromised, and did not occur

99 in wild-type plants subjected to stress or in non-stressed mutant controls (Ito et
100 al., 2011).

101 RdDM can direct DNA methylation to cytosines in all sequence contexts,
102 i.e., CG, CHG, and CHH, where H represents any nucleotide excluding G (Matzke
103 and Mosher, 2014). In plants, RdDM is the only pathway that adds *de novo* DNA
104 methylation to unmethylated regions. The canonical RdDM pathway involves two
105 main processes: the biogenesis of small interfering RNAs (siRNAs), and DNA
106 methylation at target loci in DNA (Erdmann and Picard, 2020; Matzke and
107 Mosher, 2014). Initially, the RNA Polymerase IV (Pol IV) interacts with the
108 chromatin remodeler CLASSY 1 (CLSY1) and the SAWADEE HOMEODOMAIN
109 HOMOLOG 1 (SHH1), and forms a complex that binds to heterochromatin. Pol
110 IV transcribes short single-stranded RNAs (ssRNAs), about 30 to 45 nucleotides
111 (nt) in length, which are converted to double-stranded RNAs (dsRNAs) by RNA-
112 DEPENDENT RNA POLYMERASE 2 (RDR2) associated with Pol IV. dsRNAs
113 are cleaved into 24-nt siRNAs by the endoribonuclease DICER-LIKE 3 (DCL3).
114 Then, the 24-nt siRNAs are methylated at the 3' end by HUA ENHANCER 1
115 (HEN1) and incorporated into ARGONAUTE 4 or 6 (AGO4, AGO6) proteins. The
116 AGO-sRNA duplex binds to complementary RNA transcribed by RNA
117 Polymerase V (Pol V), and recruits the DNA methyltransferase DOMAINS
118 REARRANGED METHYLTRANSFERASE 2 (DRM2), which methylates nearby
119 DNA (Erdmann and Picard, 2020; Lucibelli et al., 2022; Matzke and Mosher,
120 2014).

121 The CLSY1 protein is characterized by the presence of the SNF2 and
122 Helicase C domains, and acts together with RDR2 and NRPD1a, a subunit of Pol
123 IVa, in the production of 24-nt siRNAs (Smith et al., 2007). Besides CLSY1, the
124 CLASSY family comprises three other members called CLSY2, CLSY3, and
125 CLSY4 (Law et al., 2011). The four CLSYs are required for global and locus-
126 specific regulation of DNA methylation. In locus-specific regulation, different
127 chromatin modifications occur to produce 24-nt siRNAs, depending on which
128 CLSY proteins are involved. CLSY1 and CLSY2 are required for the association
129 between SHH1 and the Pol IV complex, in a manner dependent on histone H3
130 lysine 9 (H3K9) methylation. In the production of 24-nt siRNAs at loci controlled
131 by CLSY3 and CLSY4, CG methylation is required (Zhou et al., 2018). A study
132 has shown that, in addition to acting in the canonical RdDM pathway, the four

133 CLSY proteins mediate DNA demethylation at specific loci, demonstrating the
134 dual role that the CLSY family plays in balancing methylation and demethylation
135 reactions (Yang et al., 2018). Furthermore, other research revealed that CLSY1-
136 4 can control tissue-specific DNA methylation patterns in *Arabidopsis*. Tissues
137 with different CLSY expression levels had distinct DNA methylation patterns. For
138 example, the four CLSY genes were expressed in flower buds, CLSY3 exhibited
139 strong expression in ovules, and CLSY1 was expressed in leaves and rosettes.
140 These findings reveal that locus-specific regulation in conjunction with tissue-
141 specific expression of CLSYs generates epigenetic diversity during plant
142 development (Zhou et al., 2022). To date, the CLSY gene family has been
143 identified and characterized only in *Arabidopsis*. Here, we identify novel CLSY
144 members in different plant species and show their phylogenetic relationships. We
145 also investigate the expression of a potential CLSY gene during soybean
146 germination under abiotic stress conditions. These data will contribute to the
147 knowledge of the structural characteristics, evolutionary relationships, and
148 expression patterns of CLSYs.

149

150 **3.2.3. MATERIALS AND METHODS**

151

152 **3.2.3.1. Identification of CLSY gene family members**

153 In order to identify CLSY genes in 12 plant species, the genomic and
154 protein sequences of *Amborella trichopoda* v1.0, *Aquilegia caerulea* v3.1,
155 *Arabidopsis thaliana* TAIR10, *Brachypodium distachyon* v3.2, *Chlamydomonas*
156 *reinhardtii* v5.5, *Glycine max* Wm82.a2.v1, *Oryza sativa* v7.0, *Phaseolus vulgaris*
157 v2.1, *Physcomitrella patens* v3.3, *Sorghum bicolor* v3.1.1, *Vitis vinifera* v2.1, and
158 *Zea mays* RefGen_V4 were obtained from the Phytozome v13
159 (<https://phytozome-next.jgi.doe.gov/>) (Goodstein et al., 2012). We used the
160 profile Hidden Markov Models (profile HMMs) corresponding to the SNF2
161 (PF00176) and Helicase C (PF00271) domains available on the Pfam database
162 (<http://pfam.xfam.org/>) (Mistry et al., 2021) to search for protein sequences with
163 these profiles in *A. thaliana* proteome. Next, we executed hmmsearch (HMMER
164 package) with the parameters -T 60 and 50 for SNF2 and Hel_C profiles,
165 respectively. The multiple sequence alignments of the resulting proteins were

166 used as input for the program TABAJARA (Tool for Alignment Block Analysis
167 Joining Appropriate Rational Approaches), available at
168 <https://github.com/gruberlab/TABAJARA>, to construct specific profile HMMs for
169 CLSY proteins. This program calculates position-specific information scores in a
170 multiple sequence alignment and builds profile HMMs, identifying conserved and
171 specific regions in biological sequences (Oliveira et al., 2023; Oliveira and
172 Gruber, 2021). TABAJARA utilizes hmmsearch, hmmbuild and nhmmer from the
173 HMMER package (Eddy, 2011) and MUSCLE (Edgar, 2004). Three groups of
174 sequences containing the SNF2 and Hel_C domains served as input for
175 TABAJARA and the generated models were used with the HMM-Prospector
176 program (Oliveira et al., 2023; Oliveira and Gruber, 2021), available at
177 <https://github.com/gruberlab/hmmprospector>, to detect CLSY proteins in the plant
178 sequence datasets. The CLSY-specific profile HMMs constructed and used in
179 this work are available as Supplementary Material.

180

181 **3.2.3.2. Phylogenetic analysis**

182 Multiple alignments of protein sequences were performed using the
183 MUSCLE tool (Edgar, 2004). The phylogenetic tree was constructed with the IQ-
184 TREE software (Nguyen et al., 2015) using the maximum likelihood method with
185 1000 pseudoreplicates. The model VT+F+I+G4 was automatically evaluated by
186 ModelFinder (Kalyaanamoorthy et al., 2017) as the best-fit model. The
187 phylogenetic tree was edited and visualized on Figtree v1.4.4
188 (<http://tree.bio.ed.ac.uk/software/figtree/>).

189

190 **3.2.3.3. Gene structure and domain prediction**

191 The gene structure was generated based on the coding sequence, 5'UTR,
192 and 3'UTR of each gene in the Gene Structure Display Server (GSDS) 2.0
193 (<http://gsds.gao-lab.org/>) (Hu et al., 2015). The prediction of the domains was
194 performed in the Pfam database (<http://pfam.xfam.org/>) (Mistry et al., 2021). The
195 domain sequences were aligned using MUSCLE with default parameters in the
196 Jalview software (Waterhouse et al., 2009), version 2.11.3.2.

197

198 **3.2.3.4. Non-synonymous and synonymous (Ka/Ks) analysis for**
199 **duplicated pairs of CLASSY genes in soybean**

200 Soybean genome sequences (*Glycine max* Wm82.a2.v1) along with GFF3
201 annotation files were obtained from Phytozome v13 database (Goodstein et al.,
202 2012). The coding sequences (CDS) of CLSY genes were aligned based on the
203 amino acid sequence orientations using ClustalW with its default settings (Larkin
204 et al., 2007). Duplication events and collinearity relationships were identified
205 using the Multiple Collinearity Scan toolkit (MCScanX) (Wang et al., 2012). The
206 non-synonymous (Ka) and synonymous substitution (Ks) rate were computed
207 using KaKs_Calculator 3.0 (Zhang, 2022) using the NG method (Nei and
208 Gojobori, 1986). The duplication date (million years ago, Mya) was estimated
209 using the formula $T = Ks/2\lambda \times 10^{-6}$ Mya, ($\lambda = 6.5 \times 10^{-9}$) (Lynch and Conery, 2000).
210

211 **3.2.3.5. Expression analysis of CLSY genes in Arabidopsis and**
212 **soybean tissues**

213 The ePlant suite (<https://bar.utoronto.ca/eplant/>) (Fucile et al., 2011) was
214 used to compare the expression patterns of CLSY genes already identified in
215 *Arabidopsis*. To analyze the expression profiles of predicted CLSY genes in
216 soybean tissues, the Soybean Expression Atlas v2 database was used
217 (<https://soyatlas.venanciogroup.uenf.br/>) (Almeida-Silva et al., 2023). The
218 expression of these genes was calculated in Transcripts Per Million (TPM),
219 converted to log2, and visualized using heatmaps generated by the heatmap.2
220 function available in the gplots package in R (Keen, 2018).
221

222 **3.2.3.6. Expression analysis of a CLSY gene during soybean**
223 **germination under stress conditions**

224 Initially, the soybean seeds from cultivars BR-16 (drought-sensitive) and
225 Embrapa 48 (drought-tolerant) were disinfected with 2% of sodium hypochlorite
226 and then washed five times with sterilized distilled water. Seeds were placed in
227 Petri dishes (90 × 15 mm) with 12 ml of sterilized distilled water on gemitest
228 paper and treated with sodium chloride (NaCl) and mannitol at concentrations of
229 100 mM and 300 mM, respectively. For each treatment, two biological replicates
230 were used, each comprising a pool of 10 seeds. Treated and untreated seeds

231 were germinated in a biochemical oxygen demand (B.O.D.) incubator at 28 °C,
232 without photoperiod. With 30 hours after imbibition (HAI), the radicles were
233 manually sectioned for total RNA extraction using Trizol® (Invitrogen) according
234 to the manufacturer's instruction. The integrity of the RNA was verified by
235 electrophoresis on 1% agarose gel stained with ethidium bromide. RNA
236 quantification was carried out using a NanoDrop™ One spectrophotometer
237 (Thermo Fisher Scientific). The synthesis of cDNA was performed with 5 µg of
238 RNA using the SuperScript™ III kit (Invitrogen). The gene expression was
239 evaluated by RT-qPCR with SYBR Green PCR Master Mix (Applied Biosystems),
240 using the StepOne™ Real-Time PCR System (Applied Biosystems). The forward
241 and reverse sequences of the primers used are listed in Table S1. The METK4
242 gene was used as constitutive. Relative expression was calculated using the $2^{-\Delta\Delta CT}$
243 method (Rao et al., 2013). The data obtained were submitted to the t-Test
244 with a significance level of 95% ($P < 0.05$) in the GraphPad Prism v9.0 program
245 (<https://www.graphpad.com/>).

246

247 **3.2.4. RESULTS AND DISCUSSION**

248

249 **3.2.4.1. Construction of specific profile HMMs of CLSY proteins**

250 The identification of CLSY members in plant genomes is challenging, once
251 this family of proteins contains two widespread domains – SNF2 and Helicase C
252 (Smith et al. 2007). To identify CLSY members in plants, we constructed specific
253 profile HMMs using CLSYs from *Arabidopsis*. Initially, all 35,386 *Arabidopsis*
254 protein sequences were downloaded from Phytozome v13. To identify CLSY
255 proteins in this dataset, we used hmmsearch (HMMER package) with Pfam-
256 derived profile HMMs of SNF2 and Helicase C profile HMMs. In total, 41 proteins
257 containing the SNF2 domain and 123 proteins with the helicase C domain were
258 found. Of these, 35 proteins contained both domains (Fig. 1).

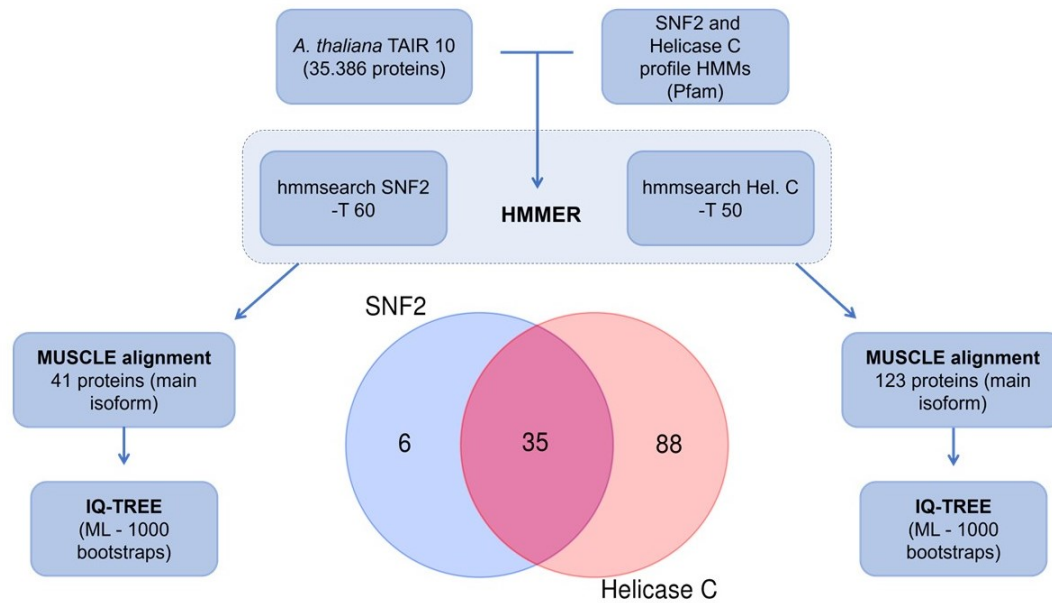


Fig. 1. Workflow of the protein search approach using profile HMMs of SNF2 and Helicase C in *Arabidopsis*, followed by phylogenetic analysis.

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

The proteins from each domain and those having both domains were aligned using MUSCLE and then utilized to generate phylogenetic trees in IQ-TREE (Fig. S1-S3). In the resulting trees, the four CLSY proteins of *Arabidopsis*, CLSY1 (AT3G42670), CLSY2 (AT5G20420), CLSY3 (AT1G05490), and CLSY4 (AT3G24340), are highlighted in red. The tree of 41 proteins containing the SNF2 domain (six proteins with solely the SNF2 domain and 35 containing both) presented a monophyletic group with the four CLSYs (Fig. S1). The tree with the 123 proteins featuring the Helicase C domain (88 proteins containing only the Helicase C domain and 35 with both) also exhibited a monophyletic group of CLSY proteins (Fig. S2). Fig S3 also displays a phylogenetic tree of 35 proteins that contain both SNF2 and Helicase C domains. Two additional trees were also generated, the former using sequences covering only the 41 SNF2 and excluding the Helicase C domain of the 35 proteins with both domains (Fig. S4), and the latter using 123 Helicase C domains, excluding the SNF2 domain of the 35 proteins (Fig. S5). The tree derived from the SNF2 domain, without the Helicase C domain (Fig. S4), showed no difference in the topology compared to the tree using both domains (Fig. S1), with the four CLSYs remaining in a monophyletic group. This result suggests that these CLSY proteins are monophyletic when analyzed only from the point of view of the SNF2 domain. However, in the tree

279 constructed exclusively with Helicase C domain sequences, the four CLSY
280 proteins were divided into three different groups, with only the CLSY1 and CLSY2
281 proteins in a single clade (Fig. S5). This indicates that CLSY proteins are
282 paraphyletic when analyzed solely using the Helicase C domain. Together, these
283 results suggest that the CLSY protein group differentiated from other proteins
284 containing the SNF2 domain and that the Helicase C domain was subsequently
285 incorporated in distinct events.

286 CLSY-specific profile HMMs were constructed with individual executions
287 of TABAJARA using protein datasets of the 41 complete proteins shown in the
288 tree of Fig. S1, 123 complete proteins represented in Fig. S2, and 41 proteins
289 containing solely the SNF2 domain (Fig. S4), respectively. The profile HMMs
290 generated from each dataset were validated for the detection of the four
291 *Arabidopsis* CLSY proteins. Models derived from each dataset were
292 concatenated and used with the HMM-Prospector program to interrogate different
293 plant protein datasets for the detection of novel CLSY proteins.

294

295 **3.2.4.2. Identification, phylogenetic relationship, and structural** 296 **analysis of CLSY family in plants**

297 The specific profile HMMs detected corresponding proteins in 11 of the 12
298 plant species analyzed. We run the HMM-Prospector with `-rc` (reduce cutoffs)
299 parameter set up to 1.0, 0.8, and 0.6 for all 12 plant genomes. Next, we compared
300 the resulting proteins in each dataset to verify the best cutoff. Using 0.6 reduced
301 cutoff, we found a total of 56 putative CLSY proteins from: moss *P. patens* (Pp –
302 1); basal angiosperm *A. trichopoda* (AmTr – 2); basal eudicot *A. coerulea* (Aqcoe
303 – 6); monocots *B. distachyon* (Bradi – 6), *O. sativa* (LOC_Os – 6), *S. bicolor*
304 (Sobic – 5), and *Z. mays* (Zm – 5); and dicots *A. thaliana* (AT – 6), *G. max* (Glyma
305 – 9), *P. vulgaris* (Phvul – 6), and *V. vinifera* (VIT – 4). The search for CLSY
306 proteins in the unicellular alga *C. reinhardtii* did not generate significant results
307 based on the criteria used. Similarly, a phylogenetic analysis showed that other
308 components of the RdDM pathway are absent in *C. reinhardtii*, such as NRPD1,
309 DRM2, and DEFECTIVE IN RNA DIRECTED DNA METHYLATION 1 (DRD1), a
310 member of the DDR complex involved in Pol V pathway (Matzke et al., 2015). A
311 study that investigated methylation patterns in plants and animals found that *C.*

312 *reinhardtii* has unusual methylation patterns, and probably has different
313 mechanisms than flowering plants (Feng et al., 2010).

314 To understand the phylogenetic relationship of CLSY family members, a
315 phylogenetic tree was generated with 447 proteins containing the SNF2 and
316 Helicase C domains from the 11 species, including the 56 putative CLSY proteins
317 (Fig. S6). We observed a monophyletic group containing three clades: clade 1,
318 with 13 proteins, including CLSY1 and CLSY2 from *Arabidopsis*, represented in
319 light blue; clade 2, with 22 proteins, containing *Arabidopsis* CLSY3 and CLSY4,
320 displayed in dark blue; clade 3, with 20 proteins, including the other two proteins
321 found in *Arabidopsis* (AT2G16390 and AT2G21450) that are not classified as
322 CLSY, represented in purple; and an outlier protein from *P. patens*
323 (Pp3c25_10710V3.1), shown in green (Fig. 2).

324 The clade 1, which includes CLSY1-2, presented nine of the 11 species
325 analyzed, with no proteins from the species *P. patens* and *A. trichopoda*. In clade
326 2, which contains CLSY3-4, 10 of the 11 plant species are present, with only *P.*
327 *patens* missing. The occurrence of CLSY family members from basal
328 angiosperm, basal eudicot, monocots, and eudicots indicates the presence of this
329 family throughout plant evolution.

330 The clade 3 possibly comprises proteins from the DRD1 family, a subfamily
331 of SNF2 chromatin-remodeling proteins, which also harbor the Helicase C
332 domain (Bargsten et al., 2013). Yang and collaborators (2018) showed that the
333 protein AT2G16390, classified as DRD1, forms a clade with AT2G21450 and with
334 the three *O. sativa* proteins (LOC_Os03g06920, LOC_Os06g14440, and
335 LOC_Os07g25390) contained in clade 3. The CLSY1-4 from *Arabidopsis* and
336 DRD1 formed a monophyletic group with proteins from the SNF2 gene family,
337 which indicates that CLSY and DRD1 proteins are closely related (Yang et al.,
338 2018). Besides, *Arabidopsis* CLSY1 is considered a homologue of DRD1 (Smith
339 et al. 2007). Taking all this results into account, it is feasible to assume that a
340 monophyletic group containing CLSY and DRD1 proteins could be formed, as
341 shown in Fig. 2.

342 A single *P. patens* protein was an outlier of the other three clades. It has
343 already been reported that some components of the RdDM pathway are absent
344 in *P. patens*, such as RNA-DIRECTED DNA METHYLATION 1 (RDM1), which is
345 part of the DDR complex as well as DRD1 (Matzke et al., 2015). The absence of

346 specialized Pol V pathway proteins, which facilitate the *de novo* methylation step,
 347 may indicate that *P. patens* has a separate pathway of Pol V evolution or
 348 represents an interrupted early stage before acquiring components existing in
 349 flowering plants (Matzke et al., 2015).

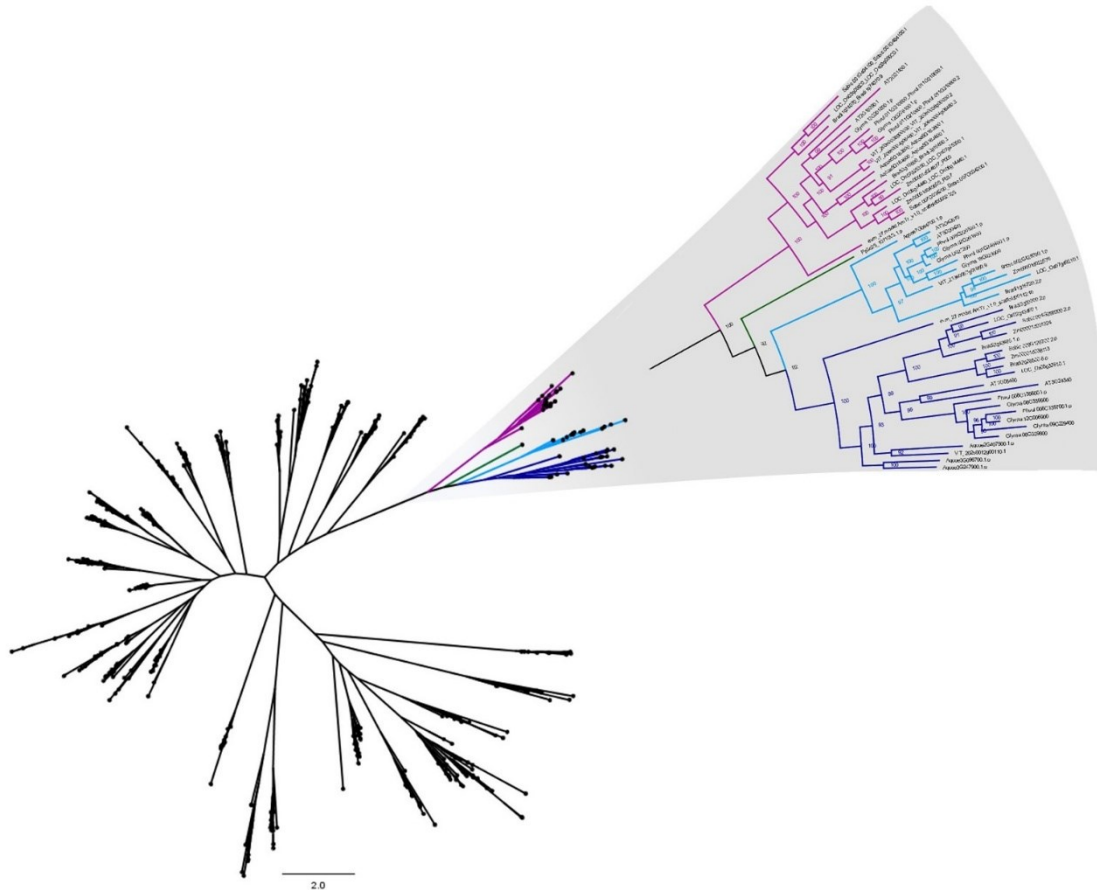


Fig. 2. Phylogenetic tree of proteins containing the SNF2 and Helicase C domains from 11 species, highlighting the monophyletic group containing the CLSY and DRD1 proteins. IQ-TREE software was used to construct the tree utilizing the maximum likelihood method. The values showed on the branches represent the bootstrap values from 1000 pseudoreplicates. Light blue, dark blue, and purple represent clades 1, 2, and 3, respectively. The single *P. patens* protein is shown in green. The complete tree can be found in the supplementary material (Fig. S6).

350

351

352 The gene ID, gene length, number of exons, protein length, presence of
 353 other domains, and chromosomal location of the identified proteins were
 354 compiled and are listed in Table 1. The length of the genes varied from 1,938 to
 355 15,347 base pairs (bp). The number of exons ranged from 1 to 12. The protein
 356 length varied from 405 to 1,875 amino acids (aa). Only three proteins presented
 357 other domains. The Glyma.18G023900 and Pp3c25_10710V3.1 proteins
 358 exhibited the SAWADEE domain, and the LOC_Os07g49210.1 protein showed
 359 the Methyltransferase domain.

360 **Table 1.** Gene identification and sequence characteristics of the 56 corresponding
 361 proteins identified from the profile HMMs.

Gene ID	Gene length (bp)	Exons number	Protein length (aa)	Other domains	Chromosome	Phylogenetic tree clade
Aqcoe7G044700.1	6524	5	1312	0	7	
AT3G42670	4288	5	1257	0	3	
AT5G20420	5153	5	1262	0	5	
Bradi1g16720.2	5594	5	1261	0	1	
Glyma.02G261800	7254	5	1311	0	2	
Glyma.18G023900	5085	5	1236	SAWADEE	18	
Glyma.U027200	7454	5	1308	0	scaffold_265	Clade 1
LOC_Os07g49210.1	10327	9	1875	Methyltransferase	7	
Phvul.001G246400.1	4915	5	1179	0	1	
Phvul.008G220500.1	5002	5	1311	0	8	
Sobic.002G428700.1	5539	5	1231	0	2	
VIT_213s0067g01950.9	5914	5	1264	0	13	
Zm00001d022576	4684	5	1335	0	7	
AmTr_v1.0_scaffold00142.46	3931	3	405	0	scaffold00142	
Aqcoe2G407500.1	3357	3	761	0	2	
Aqcoe3G096700.1	3828	3	719	0	3	
Aqcoe3G247900.1	5106	3	1173	0	3	
AT1G05490	4851	3	1411	0	1	
AT3G24340	3638	3	1133	0	3	
Bradi2g26500.6	8975	3	1507	0	2	
Bradi2g43495.1	6754	3	1286	0	2	
Bradi3g50300.2	6593	3	1416	0	3	
Glyma.08G339800	4517	3	1149	0	8	
Glyma.08G339900	7386	3	1247	0	8	Clade 2
Glyma.09G229400	3769	5	618	0	9	
Glyma.12G006900	4114	3	1167	0	12	
LOC_Os02g43460.1	5383	3	1440	0	2	
LOC_Os05g32610.1	5705	3	1446	0	5	
Phvul.008G139600.1	4598	3	1143	0	8	
Phvul.008G139700.1	4337	3	1219	0	8	
Sobic.004G299200.2	5147	4	1279	0	4	
Sobic.009G126700.2	7022	3	1459	0	9	
VIT_202s0012g00110.1	1938	1	646	0	2	
Zm00001d038113	7648	4	1436	0	6	
Zm00001d051324	4853	3	1338	0	4	
AmTr_v1.0_scaffold00002.323	15347	9	1095	0	scaffold00002	
Aqcoe5G183800.1	8214	6	1060	0	5	
Aqcoe5G184600.1	8072	6	1027	0	5	Clade 3
AT2G16390.1	3982	5	889	0	2	
AT2G21450.1	2947	4	817	0	2	

Bradi1g74070.9	5856	5	974	0	1
Bradi3g19890.3	7475	4	948	0	3
Glyma.12G236100.1	7244	6	884	0	12
Glyma.13G201800.1	6578	6	954	0	13
LOC_Os03g06920.1	7317	7	1198	0	3
LOC_Os06g14440.1	7192	6	952	0	6
LOC_Os07g25390.1	6671	5	967	0	7
Phvul.011G210600.2	6631	6	901	0	11
Phvul.011G210800.1	4894	5	872	0	11
Sobic.001G494100.1	6270	6	946	0	1
Sobic.007G034200.1	6488	6	971	0	7
VIT_203s0038g00030.2	13882	5	973	0	3
VIT_206s0004g08480.3	8920	5	976	0	6
Zm00001d024677	5604	6	951	0	10
Zm00001d049605	7240	5	978	0	4
Pp3c25_10710V3.1	8807	12	1534	SAWADEE	25

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

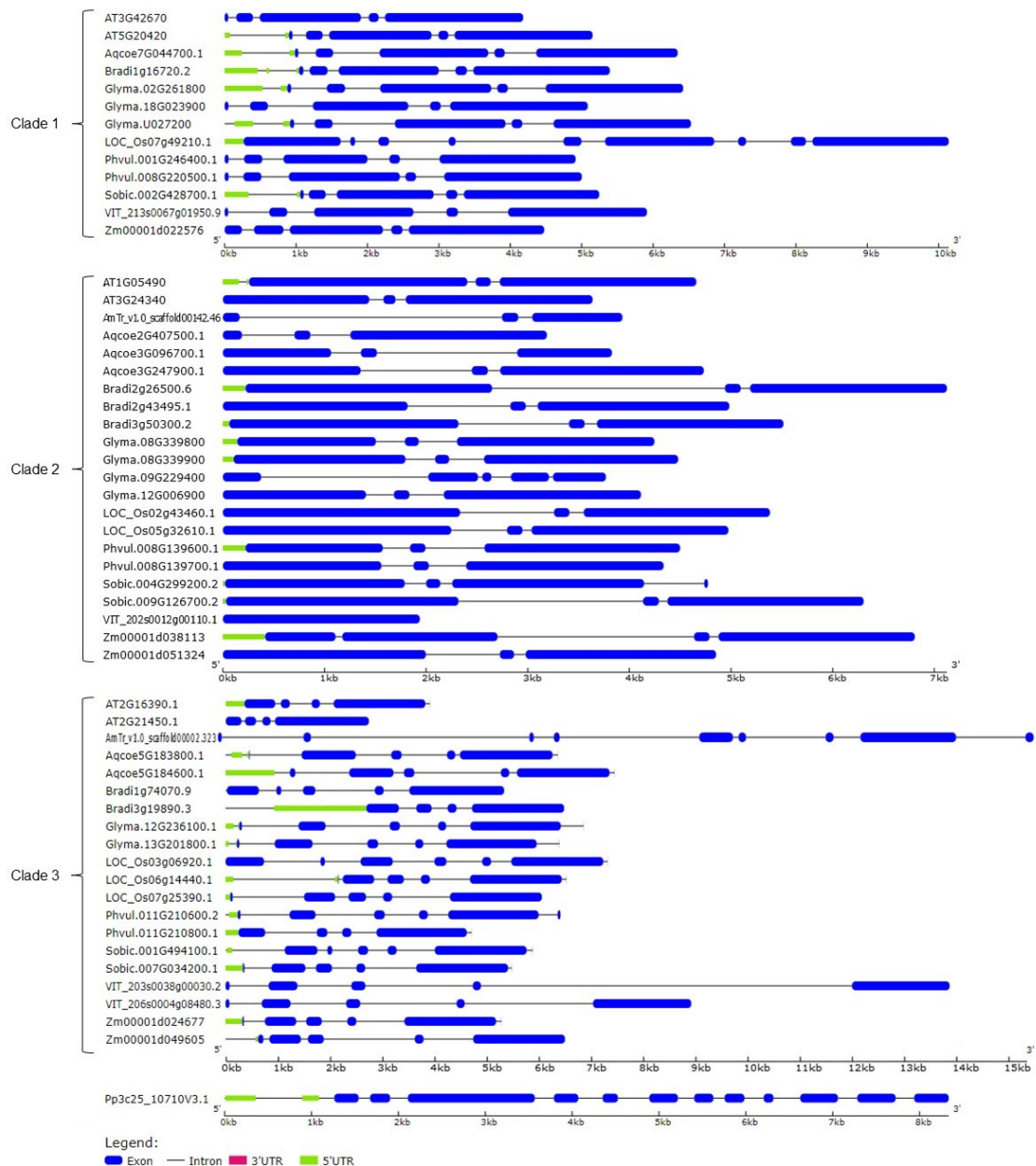
382

383

We also analyzed intron-exon patterns to understand the structural diversity of CLSY genes. As shown in Fig. 2, CLSY family members can be divided into two clades (clades 1 and 2) considering the division of *Arabidopsis* CLSY1-4. In the first clade, 12 of 13 genes had five exons, with gene lengths ranging from 4288 to 7454 bp. Only the LOC_Os07g49210.1 gene from *O. sativa* had nine exons and a size above 10 kb. In the second clade, 18 of 22 genes had three exons, and in the remaining four genes the number of exons varied from 1 to 5. The genes in clade 3, possibly related to the DRD1 family, showed a greater variation in the number of exons compared to clades 1 and 2, ranging from 4 to 9 exons. The single *P. patens* gene showed the highest number of exons among all clades, with 12 exons (Fig. 3). In general, members of the same clade presented similar exon-intron structures. This observation reinforces the division of clades shown in the phylogenetic analysis.

To understand the domain architecture of CLSY members, their protein structures were analyzed (Fig. 4). Almost all proteins presented both SNF2 and Helicase C domains, except for the proteins AmTr_v1.0_scaffold00142.46 and Aqcoe3G096700.1 from clade 2, which did not present the Helicase C domain. Three proteins showed other domains besides SNF2 and Helicase C. The proteins Glyma.18G023900 from clade 1 and Pp3c25_10710V3.1 presented the SAWADEE domain, while LOC_Os07g49210.1 from clade 1 presented the Methyltransferase domain. These two additional domains are possibly involved

384 in RdDM pathway processes. In a general context, the similarity between the
 385 gene and protein structures of members of clade 1 with CLSY1-2 of *Arabidopsis*,
 386 as well as members of clade 2 with CLSY3-4 indicates that proteins from these
 387 two clades potentially integrate the CLSY family.



388

389 **Fig. 3.** Gene structure of CLSYs, considering the clade division shown in the
 390 phylogenetic tree. Exons and introns are represented by blue boxes and black lines,
 391 respectively. The length of exons and introns is indicated in kilobases (kb).

392

393

394

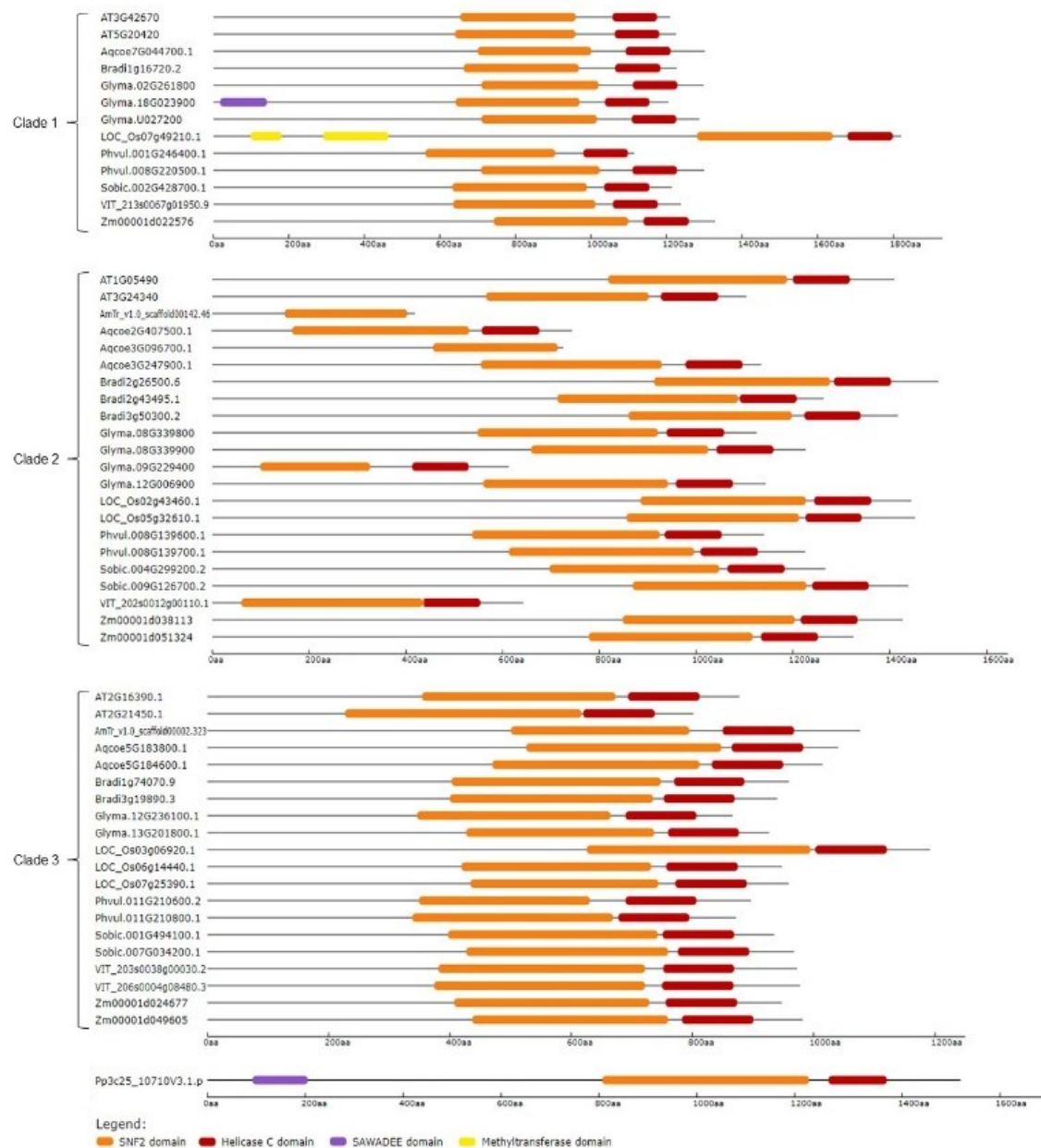
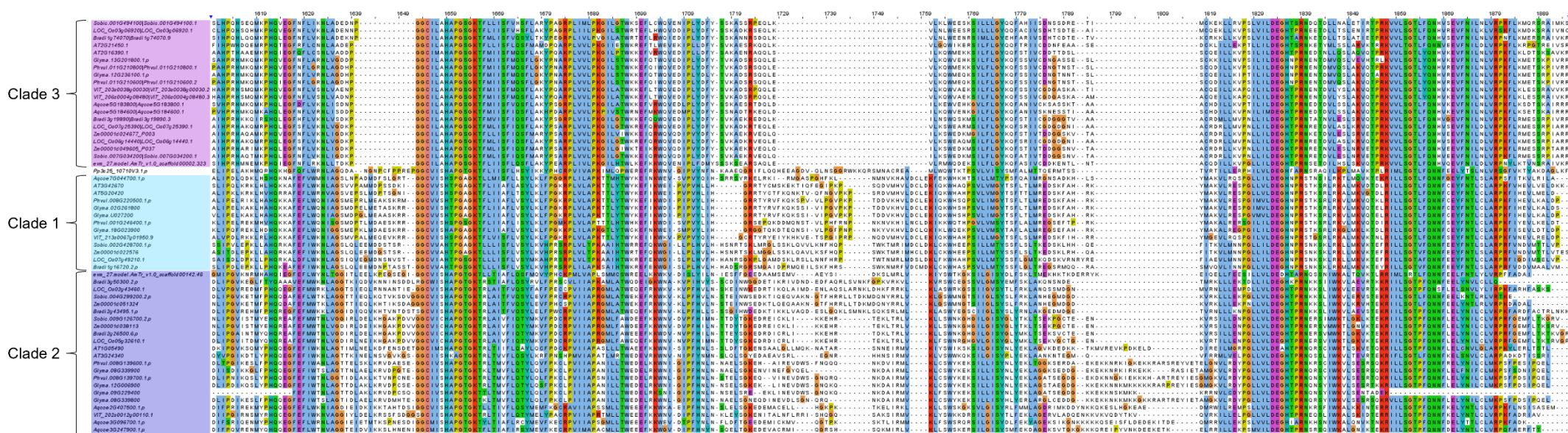


Fig. 4. Protein domain architecture following the clade division from the phylogenetic analysis. The size of proteins can be estimated from the scale in number of amino acids (aa). The corresponding color of each domain is indicated in the legend.

395

396 The multiple alignment of the SNF2 domains of the 56 CLSY proteins
 397 showed a group of amino acid residues conserved only among CLSY3-4, but
 398 absent in CLSY1-2 and the other subfamily (Fig. 5). This observation supports
 399 the result presented in the phylogenetic tree in Fig. S4, which shows that
 400 *Arabidopsis* CLSYs form a unique clade when only the SNF2 domain is used as
 401 a parameter. In contrast, the Helicase C domain is conserved among the different
 402 species in the three clades (Fig. S7), which also reinforces what was shown in

403 the phylogenetic tree containing only the Helicase C domains from *Arabidopsis*
404 (Fig. S5), where the CLSYs were not a monophyletic clade. These data
405 demonstrate the importance of the SNF2 domain for the identification of CLSY
406 proteins.



407

Fig. 5. Multiple sequence alignment of the SNF2 domain obtained in Jalview, highlighting the clades 1, 2, and 3. The clades are in the order in which they were arranged in the phylogenetic tree (Fig. S6). A color scale on amino acids was used to indicate sequence similarity.

408 **3.2.4.3. Duplication events of the CLSY genes in soybean**

409 Soybean was the species with the most genes identified in the three
410 clades, with three genes in clade 1, four genes in clade 2, and two genes in clade
411 3. To further understand the expansion of soybean genes related to the CLSY
412 family, we analyzed the duplication events of the genes included in clades 1 and
413 2, as well as the genes phylogenetically close to DRD1 in clade 3. Almost all
414 genes were duplicated through whole-genome duplication (WGD)/segmental
415 events, except Glyma.08G339900, which underwent a tandem duplication (Table
416 2). This suggests that WGD/segmental duplication played an important role in the
417 expansion of soybean genes related to the CLSY and DRD1 families. WGD
418 events correspond to complete duplication of chromosomes, and segmental
419 duplications indicate long stretches of duplicated sequences with high identity.
420 Tandem duplications produce a copy of a gene next to it, generating tandemly
421 arrayed genes (Lallemand et al., 2020). Studies have shown that, in the soybean
422 genome, the predominant type of duplication is segmental (Bolon et al., 2014;
423 Pagel et al., 2004), which is consistent with our results.

424 Considering that soybean underwent two rounds of WGD (Schmutz et al.,
425 2010), we analyzed the collinear relationships of the duplicate pairs and identified
426 ten pairs of paralogs that have close phylogenetic relationships (Table 2). All
427 soybean genes in clade 1 are paralogs to each other, and the same is true for
428 genes in clades 2 and 3. The nonsynonymous to synonymous substitutions ratio
429 values (K_a/K_s) of the genes ranged from 0.239974 to 0.754383 and were used
430 to estimate the selection pressure. The K_a/K_s ratio = 1 indicates neutral selection.
431 While a K_a/K_s ratio of < 1 corresponds to negative selection and a K_a/K_s ratio of
432 > 1 implies positive selection (Hurst, 2002). All K_a/K_s ratio values obtained were
433 < 1 , which denotes that soybean paralogous genes are under purification or
434 stabilization selection.

435 It was observed that the duplication date of the ten pairs of paralogs
436 ranged from 9.03 to 47.6 Mya (Table 2). The paralog pair
437 Glyma.02G261800/Glyma.U027200 showed the lowest K_a/K_s value (~ 0.24) and
438 the shortest divergence time (9.03 Mya), which may indicate that this gene pair
439 maintained its functions after duplication. The soybean genome underwent two
440 WGD events, the first 59 Mya ago and the second 13 Mya ago (Schmutz et al.,

441 2010). Of the 10 paralog pairs, eight were duplicated after the first WGD event
 442 and two were duplicated close to the second event. Therefore, the two duplication
 443 events may have contributed to the expansion of these genes in soybean.

444

445 **Table 2.** Paralog pairs genes, Ka/Ks ratio values, and duplication type of soybean genes.

Gene ID	Gene ID	Ka	Ks	Ka/Ks	Duplica- tion Date (Mya)	Selection pressure	Duplica- tion Type
Glyma.02G261800	Glyma.18G023900	0.223722	0.618747	0.361573	47.60	Purification or Stabilization selection	WGD or Segmental
Glyma.02G261800	Glyma.U027200	0.0281611	0.117351	0.239974	9.03	Purification or Stabilization selection	WGD or Segmental
Glyma.U027200	Glyma.18G023900	0.223407	0.614387	0.363626	47.26	Purification or Stabilization selection	WGD or Segmental
Glyma.08G339900	Glyma.08G339800	0.219666	0.373098	0.58876	28.70	Purification or Stabilization selection	Tandem
Glyma.08G339900	Glyma.09G229400	0.166574	0.407391	0.40888	31.34	Purification or Stabilization selection	WGD or Segmental
Glyma.08G339900	Glyma.12G006900	0.225172	0.396627	0.567	30.51	Purification or Stabilization selection	WGD or Segmental
Glyma.12G006900	Glyma.09G229400	0.131003	0.173655	0.754383	13.36	Purification or Stabilization selection	WGD or Segmental
Glyma.12G006900	Glyma.08G339800	0.14434	0.301214	0.47919	23.17	Purification or Stabilization selection	WGD or Segmental
Glyma.09G229400	Glyma.08G339800	0.14488	0.309337	0.468356	23.80	Purification or Stabilization selection	WGD or Segmental
Glyma.12G236100	Glyma.13G201800	0.09762	0.31009	0.3148	23.85	Purification or Stabilization selection	WGD or Segmental

446

447 **3.2.4.4. Tissue expression profile of CLSY genes in Arabidopsis and** 448 **soybean**

449 To analyze the expression profiles of the *Arabidopsis* genes identified in
 450 clades 1, 2, and 3, the tissues showing the highest expression of each gene in
 451 the ePlant database were selected. The tissues include the shoot apex, the seed

452 at 24 hours of imbibition, seed tissues at different stages of development, parts
453 of the flower, and young seeds. The genes CLSY1 and AT2G16390 (DRD1)
454 showed higher expression in the shoot apex transition and inflorescence tissues
455 compared to the other genes. Only CLSY1 presented high expression in the 24
456 h imbibed seed. In developing seed tissues, in general, all genes showed higher
457 expression in seed stages 4 to 7, except for CLSY3. The genes CLSY1, CLSY2,
458 AT2G21450, and AT2G16390/DRD1 were highly expressed in the carpel of
459 young flowers. CLSY3, CLSY4, AT2G21450 and AT2G16390/DRD1 exhibited
460 greater expression in ovules and at different stages of young seeds (Fig. S8).
461 These observations demonstrate that there are similar expression patterns
462 between the CLSYs and DRD1-related genes in different *Arabidopsis* tissues.

463 Data from different types of abiotic stresses show that the six *Arabidopsis*
464 genes analyzed here can have altered expression profiles under adverse
465 environmental conditions. The Fig. S8 lists nine types of stresses reported in
466 shoot and root. In CLSY2, for example, after 30 minutes of osmotic stress by 300
467 mM mannitol, this gene is up-regulated in the shoot. After 3 hours of the same
468 treatment, the expression of this gene is lower in the shoot compared to the
469 control (Fig. S9). This demonstrates that CLSYs expression profiles can be
470 modulated by abiotic stresses in *Arabidopsis*.

471 To investigate the expression profiles of soybean genes present in the
472 three clades, RNA-seq data from the Soybean Expression Atlas database were
473 used, relating to 14 tissues, including embryo, seed coat, seed, cotyledon, leaves,
474 callus, nodule, pod, flower, hypocotyl, suspensor, shoot, endosperm and seedling
475 (Fig. S10). The gene Glyma.U027200 showed higher expression among tissues
476 in comparison to the other genes, especially in the embryo tissue. We confirmed
477 the expression profile of this gene in embryonic axes by RT-qPCR, compared to
478 the gene Glyma.08G339900 (Fig. S11). The paralogous gene pair
479 Glyma.09G229400/Glyma.12G006900 presented a similar expression profile,
480 showing expression only in the suspensor tissue. This may be indicative that
481 these genes kept the same functions after the duplication. The other pairs of
482 paralogs showed different expression levels in one or more tissues. This
483 suggests that these genes may have acquired distinct functions after duplication.

484 We also evaluated the expression profiles of soybean genes under salt
485 and drought stress conditions by utilizing publicly available RNA-seq data (Fig.

486 6). The expression of some of the analyzed genes was induced under different
487 stress conditions. Treatment with 150 mM NaCl after six hours in seedlings
488 induced the expression of the genes Glyma.18G023900 and Glyma.12G006900
489 (Fig. 6a). In roots subjected to drought, almost all genes were induced, excluding
490 only Glyma.18G023900 and Glyma.12G006900 (Fig. 6b). Treatment with 0.9%
491 NaCl in roots induced the expression of some genes, such as: Glyma.09G229400
492 after 1, 2, and 24hours; Glyma.12G006900 and Glyma.08G339900 after 1, 2, and
493 48 hours (Fig. 6c). In shoots subjected to water deficit, the
494 Glyma.13G201800gene was induced after 12 hours, and the Glyma.08G339800
495 gene was induced after 24 hours (Fig. 6d). These data suggest that genes with
496 induced expression may play an important role in the response to salt, drought,
497 and water deficit stresses in soybean.

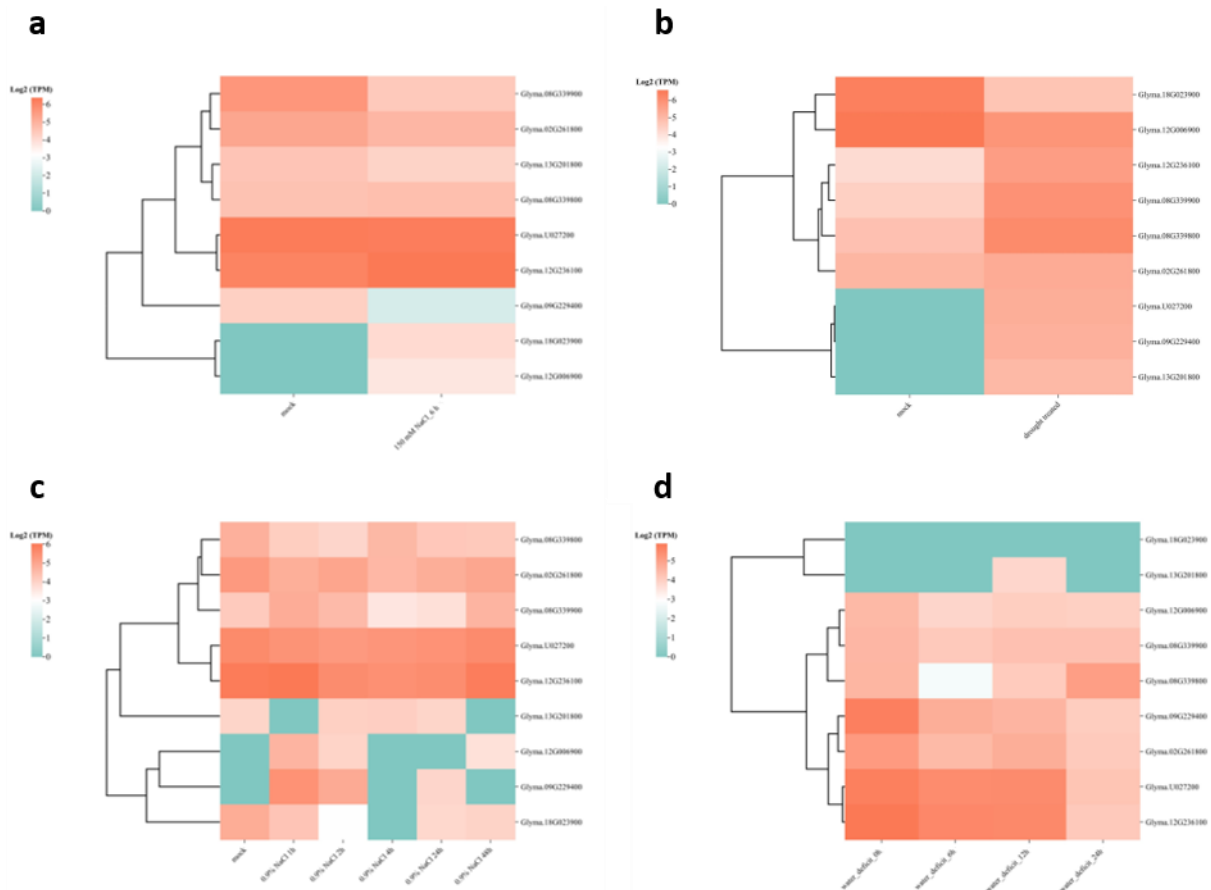


Fig. 6. Log₂ expression (TPM) of nine soybean genes from clades 1, 2, and 3 under stress conditions. Green and orange colors indicate low and high expression, respectively. (a) soybean seedlings treated with 150 mM NaCl after six hours. (b) soybean roots subjected to drought. (c) soybean roots treated with 0.9% NaCl after 1, 2, 4, 24, and 48 hours. (d) soybean shoots subjected to water deficit after 0, 6, 12, and 24 hours.

498

499 **3.2.4.5. The expression profile of CLSY and five other genes involved**
 500 **in epigenetic regulation can be modulated under abiotic stresses**
 501 **during soybean germination**

502 Considering that the Glyma.U027200 gene showed higher expression
 503 among different soybean tissues in relation to the other genes, mainly in the
 504 embryo tissue (Fig. S10), we investigated whether the expression of this gene
 505 could be altered under stress conditions in the embryonic axis during germination.
 506 For this, seeds of BR-16 (drought-sensitive) and Embrapa 48 (drought-tolerant)
 507 cultivars were treated with 100 mM NaCl and 300 mM mannitol to evaluate the
 508 effect of salt and osmotic stress on Glyma.U027200 gene expression. The
 509 embryonic axes of the seeds of both cultivars were removed after 30 hours of
 510 treatment for RNA extraction and subsequent evaluation of gene expression by

511 RT-qPCR. The Glyma.U027200 gene showed contrasting expression profiles
512 between the two cultivars, being down-regulated in BR-16 in both treatments, and
513 up-regulated in Embrapa 48, also in both stresses (Fig. 7). A study that compared
514 gene expression patterns in leaves and roots of BR-16 and Embrapa 48 under
515 water deficit also showed different genes with contrasting expression between
516 the cultivars. It was shown that Embrapa 48 presented 770 more up-regulated
517 genes than BR-16. The RNA-seq data generated in this study showed that, for
518 example, the gene Glyma.09G229400 (which appears in clade 2 of Fig. 2) has a
519 higher expression in the root of Embrapa 48 than in BR-16, in water deficit
520 treatments with 100 and 150 min (Reis et al., 2020). This indicates a possible
521 genetic/molecular difference in the responses of these cultivars to stress.

522 In addition to the analysis performed for the gene Glyma.U027200 related
523 to CLSY1-2, we also investigated the expression of five other genes involved in
524 epigenetic regulation using the same treatments (Fig. 7). The genes analyzed
525 were: DRM2, AGO4, and DCL3, which are involved in the RdDM pathway;
526 REPRESSOR OF SILENCE 1 (ROS1), responsible for removing DNA
527 methylation; and AGO1, which participates in sRNA production in the non-
528 canonical RdDM pathway (Erdmann and Picard, 2020). In cultivar BR-16, the
529 DCL3 and AGO1 genes showed no significant variation in expression compared
530 to the control under both stresses. DRM2, on the other hand, exhibited higher
531 expression in the osmotic stress, and did not show variation under salt stress.
532 ROS1 showed the highest expression in BR-16 in both stresses, especially in salt
533 stress, while AGO4 was down-regulated in both treatments. In contrast, in
534 Embrapa 48 all five genes were down-regulated in both mannitol and NaCl
535 treatments. Therefore, among the genes analyzed, only the potential CLSY
536 (Glyma.U027200) was up-regulated in Embrapa 48. As demonstrated in
537 *Arabidopsis*, CLSY genes do not appear to follow the same expression patterns
538 as other genes involved in DNA methylation. CLSY1-4 have the most diverse
539 expression patterns in comparison to other components of the RdDM pathway,
540 genes needed for the maintenance of DNA methylation, and demethylases (Zhou
541 et al., 2022). This suggested that CLSYs may act in a different timing compared
542 to other members of RdDM pathway and could have an important role in Embrapa
543 48 epigenetic response during abiotic stress or in the epigenetic background in
544 each cultivar.

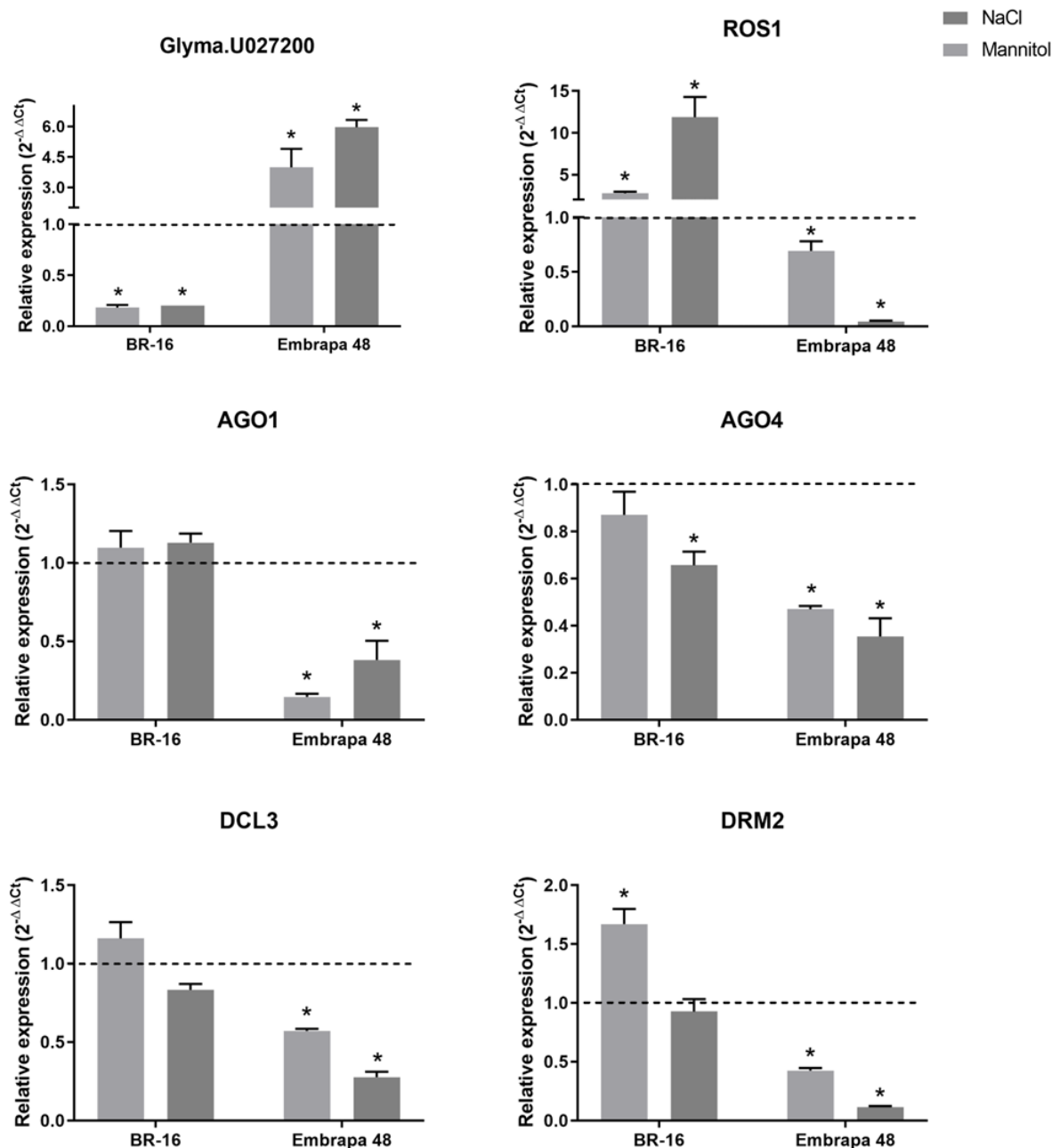


Fig. 7 Expression profiles of the AGO1, AGO4, DCL3, DRM2, Glyma.U027200, and ROS1 genes during the germination of soybean seeds treated with 300 mM mannitol and 100 mM NaCl. The dashed line represents the expression of genes in control seeds. Asterisks indicate significantly different expression values between the control and treatments, according to the t test ($P < 0.05$).

545

546 **3.2.5. CONCLUSIONS**

547

548 Data about the phylogeny and evolution of the CLSY gene family are still
 549 incipient in the scientific literature. Furthermore, there is scarce information about
 550 CLSY expression profiles under environmental stress conditions. In our study,

551 the construction of specific profile HMMs allowed the identification of new CLSY
552 proteins in 11 plant species. The proteins were divided into two clades, one
553 containing 13 proteins, including CLSY1-2 from *Arabidopsis*, and the other
554 containing 22 proteins, considering CLSY3-4 from *Arabidopsis*. From the analysis
555 of the nonsynonymous to synonymous substitution ratio values (Ka/Ks), it was
556 found that members of the CLSY family in soybean may have expanded from two
557 WGD/segmental duplication events. Considering the high expression of the
558 potential CLSY gene Glyma.U027200 in soybean embryos, we evaluated the
559 expression of this gene in germination under salt and osmotic stress in the BR-
560 16 and Embrapa 48 cultivars. The expression of the gene Glyma.U027200 and
561 five other genes involved in different epigenetic regulatory mechanisms (DRM2,
562 AGO4, DCL3, ROS1, and AGO1) showed contrasting expression profiles
563 between the two cultivars. Our work is the first to identify and analyze the
564 phylogenetic relationship of CLSY proteins in basal angiosperm, basal eudicot,
565 monocot, and eudicot species. Taken together, the results presented here
566 provide important information about the phylogeny, evolution, and gene
567 expression profiles of CLSY family members.

568

569 **Funding**

570 This work was supported by Fundação de Apoio à Pesquisa no Estado do Rio de
571 Janeiro (FAPERJ), Conselho Nacional de Desenvolvimento Científico e
572 Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível
573 Superior (CAPES) – Finance Code 001. L.S.O. received a DT scholarship from
574 CAPES.

575 **CRedit authorship contribution statement**

576 **Paula Machado de Araújo:** Formal analysis, Investigation, Writing - Original
577 Draft, Writing - Review & Editing. **Arthur Gruber:** Software, Investigation, Writing
578 - Review & Editing. **Liliane Santana Oliveira:** Software. **Sara Sangi:** Formal
579 analysis. **Geovanna Vitória Olimpio:** Formal analysis. **Felipe Cruz Paula:**
580 Formal analysis. **Clícia Grativol:** Conceptualization, Investigation, Funding
581 acquisition, Resources, Writing - Review & Editing, Supervision.

582

583 **Declaration of competing interest**

584 The authors declare that they have no known competing financial interests or
585 personal relationships that could have appeared to influence the work reported in
586 this paper.

587

588 **Appendix A. Supplementary data**

589 The following is/are the supplementary data to this article.

590

591 **Data availability**

592 Data will be made available on request.

593

594 **3.2.6. REFERENCES**

595

596 Almeida-Silva, F., Pedrosa-Silva, F., Venancio, T.M., 2023. The Soybean
597 Expression Atlas v2: A comprehensive database of over 5000 RNA-seq
598 samples. *Plant J.* 116, 1041–1051. <https://doi.org/10.1111/tpj.16459>

599 Bargsten, J.W., Folta, A., Mlynárová, L., Nap, J.P., 2013. Snf2 family gene
600 distribution in higher plant genomes reveals DRD1 expansion and
601 diversification in the tomato genome. *PLoS One* 8.
602 <https://doi.org/10.1371/journal.pone.0081147>

603 Bolon, Y.T., Stec, A.O., Michno, J.M., Roessler, J., Bhaskar, P.B., Ries, L.,
604 Dobbels, A.A., Campbell, B.W., Young, N.P., Anderson, J.E., Grant, D.M.,
605 Orf, J.H., Naeve, S.L., Muehlbauer, G.J., Vance, C.P., Stupar, R.M., 2014.
606 Genome resilience and prevalence of segmental duplications following fast
607 neutron irradiation of soybean. *Genetics* 198, 967–981.
608 <https://doi.org/10.1534/genetics.114.170340>

609 Brukhin, V., Albertini, E., 2021. Epigenetic modifications in plant development
610 and reproduction. *Epigenomes* 5.
611 <https://doi.org/10.3390/EPIGENOMES5040025>

612 Eddy, S.R., 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7.
613 <https://doi.org/10.1371/journal.pcbi.1002195>

614 Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy

- 615 and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
616 <https://doi.org/10.1093/nar/gkh340>
- 617 Erdmann, R.M., Picard, C.L., 2020. RNA-directed DNA Methylation. *PLoS Genet.*
618 16. <https://doi.org/10.1371/journal.pgen.1009034>
- 619 Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J.,
620 Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan,
621 S., Pellegrini, M., Jacobsen, S.E., 2010. Conservation and divergence of
622 methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.*
623 107, 8689–8694. <https://doi.org/10.1073/pnas.1002720107>
- 624 Fucile, G., Di Biase, D., Nahal, H., La, G., Khodabandeh, S., Chen, Y., Easley,
625 K., Christendat, D., Kelley, L., Provart, N.J., 2011. Eplant and the 3D data
626 display initiative: Integrative systems biology on the world wide web. *PLoS*
627 *One* 6. <https://doi.org/10.1371/journal.pone.0015237>
- 628 Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J.,
629 Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012.
630 Phytozome: A comparative platform for green plant genomics. *Nucleic Acids*
631 *Res.* 40, D1178–D1186. <https://doi.org/10.1093/nar/gkr944>
- 632 Grativol, C., Hemerly, A.S., Ferreira, P.C.G., 2012. Genetic and epigenetic
633 regulation of stress responses in natural plant populations. *Biochim.*
634 *Biophys. Acta - Gene Regul. Mech.* 1819, 176–185.
635 <https://doi.org/10.1016/j.bbagr.2011.08.010>
- 636 He, L., Zhao, C., Zhang, Q., Zinta, G., Wang, D., Lozano-Durán, R., Zhu, J.K.,
637 2021. Pathway conversion enables a double-lock mechanism to maintain
638 DNA methylation and genome stability. *Proc. Natl. Acad. Sci. U. S. A.* 118.
639 <https://doi.org/10.1073/pnas.2107320118>
- 640 Hemenway, E.A., Gehring, M., 2023. Epigenetic Regulation During Plant
641 Development and the Capacity for Epigenetic Memory. *Annu. Rev. Plant*
642 *Biol.* 74, 87–109. <https://doi.org/10.1146/annurev-arplant-070122-025047>
- 643 Henderson, I.R., Jacobsen, S.E., 2007. Epigenetic inheritance in plants. *Nature*
644 447, 418–424. <https://doi.org/10.1038/nature05917>

- 645 Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J., Gao, G., 2015. GSDS 2.0: An
646 upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297.
647 <https://doi.org/10.1093/bioinformatics/btu817>
- 648 Hurst, L.D., 2002. The Ka/Ks ratio: Diagnosing the form of sequence evolution.
649 *Trends Genet.* 18, 486–487. [https://doi.org/10.1016/S0168-9525\(02\)02722-](https://doi.org/10.1016/S0168-9525(02)02722-1)
650 1
- 651 Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., Paszkowski, J., 2011.
652 An siRNA pathway prevents transgenerational retrotransposition in plants
653 subjected to stress. *Nature* 472, 115–120.
654 <https://doi.org/10.1038/nature09861>
- 655 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermini, L.S.,
656 2017. ModelFinder: Fast model selection for accurate phylogenetic
657 estimates. *Nat. Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>
- 658 Keen, K.J., 2018. *Graphics for Statistics and Data Analysis with R*, 2nd ed.
659 Chapman and Hall/CRC, New York.
660 <https://doi.org/https://doi.org/10.1201/9780429031069>
- 661 Kumar, S., Mohapatra, T., 2021. Dynamics of DNA Methylation and Its Functions
662 in Plant Growth and Development. *Front. Plant Sci.* 12.
663 <https://doi.org/10.3389/fpls.2021.596236>
- 664 Kundariya, H., Sanchez, R., Yang, X., Hafner, A., Mackenzie, S.A., 2022.
665 Methylome decoding of RdDM-mediated reprogramming effects in the
666 *Arabidopsis* MSH1 system. *Genome Biol.* 23.
667 <https://doi.org/10.1186/s13059-022-02731-w>
- 668 Lallemand, T., Leduc, M., Landès, C., Rizzon, C., Lerat, E., 2020. An overview of
669 duplicated gene detection methods: Why the duplication mechanism has to
670 be accounted for in their choice. *Genes (Basel).* 11, 1–40.
671 <https://doi.org/10.3390/genes11091046>
- 672 Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A.,
673 McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson,
674 J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0.
675 *Bioinformatics* 23, 2947–2948.

- 676 <https://doi.org/10.1093/bioinformatics/btm404>
- 677 Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., Jacobsen, S.E., 2011. SHH1, a
678 Homeodomain protein required for DNA Methylation, as well as RDR2,
679 RDM4, and Chromatin remodeling factors, associate with RNA Polymerase
680 IV. *PLoS Genet.* 7. <https://doi.org/10.1371/journal.pgen.1002195>
- 681 Lucibelli, F., Valoroso, M.C., Aceto, S., 2022. Plant DNA Methylation: An
682 Epigenetic Mark in Development, Environmental Interactions, and Evolution.
683 *Int. J. Mol. Sci.* 23. <https://doi.org/10.3390/ijms23158299>
- 684 Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of
685 duplicate genes. *Science* (80-.). 290, 1151–1155.
686 <https://doi.org/10.1126/science.290.5494.1151>
- 687 Matzke, M.A., Kanno, T., Matzke, A.J.M., 2015. RNA-directed DNA methylation:
688 The evolution of a complex epigenetic pathway in flowering plants. *Annu.*
689 *Rev. Plant Biol.* 66, 243–267. [https://doi.org/10.1146/annurev-arplant-](https://doi.org/10.1146/annurev-arplant-043014-114633)
690 [043014-114633](https://doi.org/10.1146/annurev-arplant-043014-114633)
- 691 Matzke, M.A., Mosher, R.A., 2014. RNA-directed DNA methylation: An epigenetic
692 pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408.
693 <https://doi.org/10.1038/nrg3683>
- 694 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A.,
695 Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J.,
696 Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021.
697 *Nucleic Acids Res.* 49, D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- 698 Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of
699 synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*
700 3, 418–426. <https://doi.org/10.1093/oxfordjournals.molbev.a040410>
- 701 Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A
702 fast and effective stochastic algorithm for estimating maximum-likelihood
703 phylogenies. *Mol. Biol. Evol.* 32, 268–274.
704 <https://doi.org/10.1093/molbev/msu300>
- 705 Oliveira, L.S., Gruber, A., 2021. Rational Design of Profile Hidden Markov Models

- 706 for Viral Classification and Discovery, in: Nakaya, H.I. (Ed.), Bioinformatics.
707 Exon Publications, Brisbane, Australia, pp. 151–170.
708 <https://doi.org/10.36255/exonpublications.bioinformatics.2021.ch9>
- 709 Oliveira, L.S., Reyes, A., Dutilh, B.E., Gruber, A., 2023. Rational Design of Profile
710 HMMs for Sensitive and Specific Sequence Detection with Case Studies
711 Applied to Viruses, Bacteriophages, and Casposons. *Viruses* 15.
712 <https://doi.org/10.3390/v15020519>
- 713 Pagel, J., Walling, J.G., Young, N.D., Shoemaker, R.C., Jackson, S.A., 2004.
714 Segmental duplications within the Glycine max genome revealed by
715 fluorescence in situ hybridization of bacterial artificial chromosomes.
716 *Genome* 47, 764–768. <https://doi.org/10.1139/G04-025>
- 717 Pikaard, C.S., Scheid, O.M., 2014. Epigenetic regulation in plants. *Cold Spring*
718 *Harb. Perspect. Biol.* 6. <https://doi.org/10.1101/cshperspect.a019315>
- 719 Popova, O. V., Dinh, H.Q., Aufsatz, W., Jonak, C., 2013. The RdDM pathway is
720 required for basal heat tolerance in arabidopsis. *Mol. Plant* 6, 396–410.
721 <https://doi.org/10.1093/mp/sst023>
- 722 Rao, X., Huang, X., Zhou, Z., Lin, X., 2013. An improvement of the $2^{-(\Delta\Delta CT)}$
723 method for quantitative real-time polymerase chain reaction data
724 analysis. *Biostat. Bioinforma. Biomath.* 3, 71–85.
- 725 Reis, R.R., Mertz-Henning, L.M., Marcolino-Gomes, J., Rodrigues, F.A.,
726 Rockenbach-Marin, S., Fuganti-Pagliarini, R., Koltun, A., Gonçalves, L.S.A.,
727 Nepomuceno, A.L., 2020. Differential gene expression in response to water
728 deficit in leaf and root tissues of soybean genotypes with contrasting
729 tolerance profiles. *Genet. Mol. Biol.* 43, 1–17. <https://doi.org/10.1590/1678-4685-GMB-2018-0290>
- 731 Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten,
732 D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu,
733 Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan,
734 B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K.,
735 Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T.,
736 Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing,

737 R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G.,
738 Shoemaker, R.C., Jackson, S.A., 2010. Genome sequence of the
739 palaeopolyploid soybean. *Nature* 463, 178–183.
740 <https://doi.org/10.1038/nature08670>

741 Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang,
742 C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L.,
743 Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C.,
744 Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton,
745 M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W.,
746 Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E.,
747 Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N.,
748 Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J.,
749 Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K.,
750 Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J.,
751 He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko,
752 A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley,
753 E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J.,
754 Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L.,
755 Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania,
756 A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van
757 Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y.,
758 Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P.,
759 Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y.,
760 Jeddloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel,
761 D.H., McCann, M.C., Sanmiguel, P., Myers, A.M., Nettleton, D., Nguyen, J.,
762 Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A.,
763 Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M.,
764 Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu,
765 Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G.,
766 Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R.,
767 Wing, R.A., Wilson, R.K., 2009. The B73 maize genome: Complexity,
768 diversity, and dynamics. *Science* (80-.). 326, 1112–1115.
769 <https://doi.org/10.1126/science.1178534>

- 770 Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard,
771 C.S., Baulcombe, D.C., 2007. An SNF2 protein associated with nuclear RNA
772 silencing and the spread of a silencing signal between cells in Arabidopsis.
773 *Plant Cell* 19, 1507–1521. <https://doi.org/10.1105/tpc.107.051540>
- 774 Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H.,
775 Marler, B., Guo, H., Kissinger, J.C., Paterson, A.H., 2012. MCSanX: A
776 toolkit for detection and evolutionary analysis of gene synteny and
777 collinearity. *Nucleic Acids Res.* 40. <https://doi.org/10.1093/nar/gkr1293>
- 778 Wassenegger, M., Heimes, S., Riedel, L., Sanger, H.L., 1994. RNA-directed de
779 novo methylation of genomic sequences in plants. *Cell* 76, 567–576.
780 [https://doi.org/10.1016/0092-8674\(94\)90119-8](https://doi.org/10.1016/0092-8674(94)90119-8)
- 781 Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009.
782 Jalview Version 2-A multiple sequence alignment editor and analysis
783 workbench. *Bioinformatics* 25, 1189–1191.
784 <https://doi.org/10.1093/bioinformatics/btp033>
- 785 Xu, R., Wang, Y., Zheng, H., Lu, W., Wu, C., Huang, J., Yan, K., Yang, G., Zheng,
786 C., 2015. Salt-induced transcription factor MYB74 is regulated by the RNA-
787 directed DNA methylation pathway in Arabidopsis. *J. Exp. Bot.* 66, 5997–
788 6008. <https://doi.org/10.1093/jxb/erv312>
- 789 Yang, D.L., Zhang, G., Wang, L., Li, J., Xu, D., Di, C., Tang, K., Yang, L., Zeng,
790 L., Miki, D., Duan, C.G., Zhang, H., Zhu, J.K., 2018. Four putative
791 SWI2/SNF2 chromatin remodelers have dual roles in regulating DNA
792 methylation in Arabidopsis. *Cell Discov.* 4. <https://doi.org/10.1038/s41421-018-0056-8>
- 794 Zhang, H., Lang, Z., Zhu, J.K., 2018. Dynamics and function of DNA methylation
795 in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506.
796 <https://doi.org/10.1038/s41580-018-0016-z>
- 797 Zhang, Z., 2022. KaKs_Calculator 3.0: Calculating Selective Pressure on Coding
798 and Non-coding Sequences. *Genomics, Proteomics Bioinforma.* 20, 536–
799 540. <https://doi.org/10.1016/j.gpb.2021.12.002>
- 800 Zhou, M., Coruh, C., Xu, G., Martins, L.M., Bourbousse, C., Lambomez, A., Law,

801 J.A., 2022. The CLASSY family controls tissue-specific DNA methylation
802 patterns in Arabidopsis. Nat. Commun. 13. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-021-27690-x)
803 021-27690-x

804 Zhou, M., Palanca, A.M.S., Law, J.A., 2018. Locus-specific control of the de novo
805 DNA methylation pathway in Arabidopsis by the CLASSY family. Nat. Genet.
806 50, 865–873. <https://doi.org/10.1038/s41588-018-0115-y>

807

4. CONSIDERAÇÕES FINAIS

Os resultados dessa pesquisa fornecem uma visão abrangente sobre a regulação de miRNAs durante a germinação de soja sob estresse osmótico e salino. A identificação de miRNAs diferencialmente expressos entre as cultivares BR-16 e Embrapa 48 nos dois estresses, demonstra o ajuste fino que os miRNAs desempenham na regulação da expressão gênica durante a germinação. A aplicação de um miRNA exógeno utilizando uma abordagem alternativa fornece dados iniciais para futuras pesquisas, visando contribuir para o aprimoramento dessa técnica. Os resultados desse trabalho também trazem novas informações sobre a evolução e conservação de proteínas CLSY em plantas. Pela primeira vez, foram identificados membros da família CLSY em outras espécies que não a planta modelo *Arabidopsis*. Também foi demonstrado que um gene CLSY em soja possui responsividade a estresses abióticos durante a germinação. Esperamos que esses achados abram caminho para novas pesquisas sobre a função das proteínas CLSY em plantas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, E. J. et al. Soybean [*Glycine max* (L.) Merr.] breeding: History, improvement, production and future opportunities. In: AL-KHAYRI, J. M.; JAIN, S. M.; JOHNSON, D. V. (Eds.). . **Advances in Plant Breeding Strategies: Legumes**. [s.l.] Springer Nature, 2019. v. 7p. 431–516.

ARAÚJO, P. M. DE; GRATIVOL, C. In silico identification of candidate miRNA-encoded Peptides in four Fabaceae species. **Computational Biology and Chemistry**, v. 97, 2022.

BAO, N.; LYE, K. W.; BARTON, M. K. MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. **Developmental Cell**, v. 7, n. 5, p. 653–662, 2004.

BEJ, S.; BASAK, J. MicroRNAs: The Potential Biomarkers in Plant Stress Response. **American Journal of Plant Sciences**, v. 05, n. 05, p. 748–759, 2014.

BEWLEY, J. D. et al. **Seeds: Physiology of development, germination and**

dormancy, 3rd edition. 3. ed. New York, NY: Springer, 2013.

BRUKHIN, V.; ALBERTINI, E. Epigenetic modifications in plant development and reproduction. **Epigenomes**, v. 5, n. 4, 2021.

BUHTZ, A. et al. Identification and characterization of small RNAs from the phloem of *Brassica napus*. **Plant Journal**, v. 53, n. 5, p. 739–749, 2008.

CARRERA-CASTAÑO, G. et al. An updated overview on the regulation of seed germination. **Plants**, v. 9, n. 6, p. 1–42, 2020.

CATTELAN, A. J.; DALL'AGNOL, A. The rapid soybean growth in Brazil. **OCL - Oilseeds and fats, Crops and Lipids**, v. 25, n. 1, 2018.

CHEN, Q. J. et al. A miRNA-encoded small peptide, vvi-miPEP171d1, regulates adventitious root formation. **Plant Physiology**, v. 183, n. 2, p. 656–670, 2020.

CHEN, Q. JU et al. vvi-miPEP172b and vvi-miPEP3635b increase cold tolerance of grapevine by regulating the corresponding MIRNA genes. **Plant Science**, v. 325, 2022.

CONAB. **Companhia Nacional de Abastecimento - Acompanhamento da safra brasileira de grãos.** Brasília, 2024. Disponível em: <https://www.conab.gov.br/component/k2/item/download/51274_e40f1bba791d27a4c67a29c5f29781ff>

COUZIGOU, J. M. et al. Use of microRNA-encoded peptide miPEP172c to stimulate nodulation in soybean. **New Phytologist**, v. 211, n. 2, p. 379–381, 2016.

CRISP, P. A. et al. Reconsidering plant memory: Intersections between stress recovery, RNA turnover, and epigenetics. **Science Advances**, v. 2, n. 2, 2016.

D'ARIO, M.; GRIFFITHS-JONES, S.; KIM, M. Small RNAs: Big Impact on Plant Development. **Trends in Plant Science**, v. 22, n. 12, p. 1056–1068, 2017.

DHAUBHADEL, S.; MARSOLAIS, F. Transcriptomics of legume seed: Soybean a model grain legume. In: AGRAWAL, G. K.; RAKWAL, R. (Eds.). **Seed Development: OMICS Technologies Toward Improvement of Seed Quality and Crop Yield: OMICS in Seed Biology.** [s.l.] Springer, 2012. p. 129–142.

ERDMANN, R. M.; PICARD, C. L. RNA-directed DNA Methylation. **PLoS Genetics**, v. 16, n. 10, 2020.

EROKHINA, T. N. et al. Activity of Chemically Synthesized Peptide Encoded by the miR156A Precursor and Conserved in the Brassicaceae Family Plants. **Biochemistry (Moscow)**, v. 86, n. 5, p. 551–562, 2021.

EROKHINA, T. N. et al. Regulatory miPEP Open Reading Frames Contained in the Primary Transcripts of microRNAs. **International Journal of Molecular Sciences**, v. 24, n. 3, 2023.

FAO. **Food and Agriculture Organization - Commodities by country**. Disponível em: <https://www.fao.org/faostat/en/#rankings/commodities_by_country>. Acesso em: 5 jan. 2024.

FAROOQ, M. S. et al. Uncovering the Research Gaps to Alleviate the Negative Impacts of Climate Change on Food Security: A Review. **Frontiers in Plant Science**, v. 13, 2022.

FENG, S.; JACOBSEN, S. E. Epigenetic modifications in plants: An evolutionary perspective. **Current Opinion in Plant Biology**, v. 14, n. 2, p. 179–186, 2011.

GALLEGO-BARTOLOMÉ, J. DNA methylation in plants: mechanisms and tools for targeted manipulation. **New Phytologist**, v. 227, n. 1, p. 38–44, 2020.

GAO, H. et al. Transcriptomic comparison reveals genetic variation potentially underlying seed developmental evolution of soybeans. **Journal of Experimental Botany**, v. 69, n. 21, p. 5089–5104, 2018.

HENDERSON, I. R.; JACOBSEN, S. E. Epigenetic inheritance in plants. **Nature**, v. 447, n. 7143, p. 418–424, 2007.

HYMOWITZ, T. On the domestication of soybean. **Economic Botany**, v. 24, n. 4, p. 408–421, 1970.

IWASAKI, M.; PASZKOWSKI, J. Epigenetic memory in plants. **The EMBO Journal**, v. 33, n. 18, p. 1987–1998, 2014.

JONES-RHOADES, M. W.; BARTEL, D. P.; BARTEL, B. MicroRNAs and their

regulatory roles in plants. **Annual Review of Plant Biology**, v. 57, p. 19–53, 2006.

KHRAIWESH, B.; ZHU, J. K.; ZHU, J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. **Biochimica et Biophysica Acta - Gene Regulatory Mechanisms**, v. 1819, n. 2, p. 137–148, 2012.

KIDNER, C. A.; MARTIENSSEN, R. A. The developmental role of microRNA in plants. **Current Opinion in Plant Biology**, v. 8, n. 1, p. 38–44, 2005.

KUMAR, V. et al. Plant small RNAs: the essential epigenetic regulators of gene expression for salt-stress responses and tolerance. **Plant Cell Reports**, v. 37, n. 1, p. 61–75, 2018.

KUMARI, P. et al. Unravelling the Role of Epigenetic Modifications in Development and Reproduction of Angiosperms: A Critical Appraisal. **Frontiers in Genetics**, v. 13, 2022.

LAURESSERGUES, D. et al. Primary transcripts of microRNAs encode regulatory peptides. **Nature**, v. 520, n. 7545, p. 90–93, 2015.

LAURESSERGUES, D. et al. Characterization of plant microRNA-encoded peptides (miPEPs) reveals molecular mechanisms from the translation to activity and specificity. **Cell Reports**, v. 38, n. 6, 2022.

LAW, J. A.; JACOBSEN, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. **Nature Reviews Genetics**, v. 11, n. 3, p. 204–220, 2010.

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

LEI, M. et al. Regulatory link between DNA methylation and active demethylation in Arabidopsis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 11, p. 3553–3557, 2015.

LI, Y.; KUMAR, S.; QIAN, W. Active DNA demethylation: mechanism and role in plant development. **Plant Cell Reports**, v. 37, n. 1, p. 77–85, 2018.

LIU, H. H. et al. Microarray-based analysis of stress-regulated microRNAs in

Arabidopsis thaliana. **Rna**, v. 14, n. 5, p. 836–843, 2008.

LLAVE, C. et al. Endogenous and silencing-associated small RNAs in plants. **Plant Cell**, v. 14, n. 7, p. 1605–1619, 2002.

LUCIBELLI, F.; VALOROSO, M. C.; ACETO, S. Plant DNA Methylation: An Epigenetic Mark in Development, Environmental Interactions, and Evolution. **International Journal of Molecular Sciences**, v. 23, n. 15, 2022.

LUJÁN-SOTO, E.; DINKOVA, T. D. Time to wake up: Epigenetic and small-RNA-mediated regulation during seed germination. **Plants**, v. 10, n. 2, p. 1–19, 2021.

MAIZEL, A. et al. To move or not to move: roles and specificity of plant RNA mobility. **Current Opinion in Plant Biology**, v. 57, p. 52–60, 2020.

MATZKE, M. A.; MOSHER, R. A. RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. **Nature Reviews Genetics**, v. 15, n. 6, p. 394–408, 2014.

MEDIC, J.; ATKINSON, C.; HURBURGH, C. R. Current knowledge in soybean composition. **Journal of the American Oil Chemists' Society**, v. 91, n. 3, p. 363–384, 2014.

MIRYEGANEH, M.; SAZE, H. Epigenetic inheritance and plant evolution. **Population Ecology**, v. 62, n. 1, p. 17–27, 2020.

MOLNÁR, A. et al. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. **Nature**, v. 447, n. 7148, p. 1126–1129, 2007.

NONOGAKI, H.; BASSEL, G. W.; BEWLEY, J. D. Germination-still a mystery. **Plant Science**, v. 179, n. 6, p. 574–581, 2010.

ORMANCEY, M. et al. Use of microRNA-encoded peptides to improve agronomic traits. **Plant Biotechnology Journal**, v. 19, n. 9, p. 1687–1689, 2021.

ORMANCEY, M. et al. The Essentials on microRNA-Encoded Peptides from Plants to Animals. **Biomolecules**, v. 13, n. 2, 2023.

PATIL, S. et al. MicroRNA-mediated bioengineering for climate-resilience in crops. **Bioengineered**, v. 12, n. 2, p. 10430–10456, 2021.

PIKAARD, C. S.; SCHEID, O. M. Epigenetic regulation in plants. **Cold Spring**

Harbor Perspectives in Biology, v. 6, n. 12, 2014.

RAM, M. K.; MUKHERJEE, K.; PANDEY, D. M. Identification of miRNA, their targets and miPEPs in peanut (*Arachis hypogaea* L.). **Computational Biology and Chemistry**, v. 83, 2019.

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

RAPP, R. A.; WENDEL, J. F. Epigenetics and plant evolution. **New Phytologist**, v. 168, n. 1, p. 81–91, 2005.

RAY, D. K. et al. Climate variation explains a third of global crop yield variability. **Nature Communications**, v. 6, 2015.

RAZA, A. et al. miRNAs for crop improvement. **Plant Physiology and Biochemistry**, v. 201, 2023.

RAŽNÁ, K.; ŽIAROVSKÁ, J.; GÁLOVÁ, Z. MicroRNA-Based Markers in Plant Genome Response to Abiotic Stress and Their Application in Plant Genotyping. In: TUTAR, L. (Ed.). . **Non-Coding RNAs**. [s.l.] Intechopen, 2019.

SCHMUTZ, J. et al. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178–183, 2010.

SIMON, S. A.; MEYERS, B. C. Small RNA-mediated epigenetic modifications in plants. **Current Opinion in Plant Biology**, v. 14, n. 2, p. 148–155, 2011.

STIEF, A. et al. Arabidopsis miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. **Plant Cell**, v. 26, n. 4, p. 1792–1807, 2014.

SUN, L.; YUAN, Z. Seed morphology of soybean. In: LAM, H.-M.; LI, M.-W. (Eds.). . **Advances in Botanical Research**. [s.l.] Academic Press, 2022. v. 102p. 349–375.

SUNKAR, R.; LI, Y. F.; JAGADEESWARAN, G. Functions of microRNAs in plant stress responses. **Trends in Plant Science**, v. 17, n. 4, p. 196–203, 2012.

TAYLOR, R. S. et al. Evolutionary history of plant microRNAs. **Trends in Plant Science**, v. 19, n. 3, p. 175–182, 2014.

USDA. **United States Department of Agriculture - Soybean World Production.** Disponível em:

<<https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2222000>>. Acesso em: 5 jan. 2024.

VAUCHERET, H. Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. **Genes and Development**, v. 20, n. 7, p. 759–771, 2006.

WADDINGTON, C. H. The epigenotype. 1942. **International Journal of Epidemiology**, v. 41, n. 1, p. 10–13, 2012.

WILLIAMS, B. P. et al. Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat. **PLoS Genetics**, v. 11, n. 3, 2015.

WU, L. et al. DNA Methylation Mediated by a MicroRNA Pathway. **Molecular Cell**, v. 38, n. 3, p. 465–475, 2010.

XU, H. et al. Progresses, Challenges, and Prospects of Genome Editing in Soybean (*Glycine max*). **Frontiers in Plant Science**, v. 11, 2020.

YADAV, A. et al. An overview on miRNA-encoded peptides in plant biology research. **Genomics**, v. 113, n. 4, p. 2385–2391, 2021.

YU, Y.; JIA, T.; CHEN, X. The ‘how’ and ‘where’ of plant microRNAs. **New Phytologist**, v. 216, n. 4, p. 1002–1017, 2017.

ZHANG, B.; WANG, Q. MicroRNA, a new target for engineering new crop cultivars. **Bioengineered**, v. 7, n. 1, p. 7–10, 2016.

ZHANG, H.; LANG, Z.; ZHU, J. K. Dynamics and function of DNA methylation in plants. **Nature Reviews Molecular Cell Biology**, v. 19, n. 8, p. 489–506, 2018.

ZHAO, T.; ZHAN, Z.; JIANG, D. Histone modifications and their regulatory roles in plant development and environmental memory. **Journal of Genetics and Genomics**, v. 46, n. 10, p. 467–476, 2019.

ZHU, J. K. Active DNA demethylation mediated by DNA glycosylases. **Annual Review of Genetics**, v. 43, p. 143–166, 2009.

APÊNDICE

CAPÍTULO 1

Tabela S1: Sequências de primers utilizados nas análises de RT-qPCR.

Primer	Sequência
miR156-Forward	GCGGCGTGACAGAAGAGAGT
miR156-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGTGCTC
miR159-Forward	GCGGCGTTTGGATTGAAGGGA
miR159-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTAGAGC
miR160-Forward	TTCCTTGCCTGGCTCCCTGT
miR160-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTGGCAT
miR166-Forward	TTCCTTCGGACCAGGCTTCA
miR166-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGGGGAA
miR167-Forward	TTCCTTGAAGCTGCCAGCA
miR167-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACAGATCA
miR168-Forward	TTCCTTCGCTTGGTGCAGGT
miR168-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTTCCCG
miR171-Forward	TCGCGTATTGGCCTGGTTCA
miR171-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTCTGAG
miR319-Forward	TCGCGTTGGACTGAAGGGA
miR319-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGGGAGC
miR390-Forward	TCGCTAAGCTCAGGAGGGAT
miR390-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGGCGCT
miR394-Forward	TCGCGTTGGCATTCTGTCC
miR394-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGGAGGT

miR398-Forward	TCGCTTGTGTTCTCAGGTCG
miR398-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACCAGGGG
miR408-Forward	TCGCTATGCACTGCCTCTTC
miR408-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGCCAGG
miR482-Forward	TCGCTTCTTCCCTACACCTCC
miR482-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACGGTATG
miR1507-Forward	GCGGCGTCTCATTCCATACATC
miR1507-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTCAGAC
miR3522-Forward	TCGCGTGAGACCAAATGAGC
miR3522-RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTTCGCACTGGATACGACTCAGCT
Universal Reverse Primer	GTGCAGGGTCCGAGGT
AGO1_Glyma.16G217300_Forward	ACCCGGTCTGTTTCAATCG
AGO1_Glyma.16G217300_Reverse	AGCACCACTTGTCATAGAGC
AGO1_Glyma.09G167100_Forward	CAAAGCGTCACCACACAAG
AGO1_Glyma.09G167100_Reverse	AGTCAAATTCGGTGGGATGG
METK4_Forward	CTGAGCCTTTGTCTGTGTTTG
METK4_Reverse	CCTGTTATTTCCACCCCTCTTG

Tabela S2: Genes-alvo preditos para o miR168.

Gene alvo	E-value	Início do gene	Fim do gene	Fragmento do miRNA alinhado	Alinhamento	Fragmento do alvo alinhado	Inibição
Glyma.10G096300.2	2.5	863	883	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUCUCGACCUGAACCAAGUGA	Tradução
Glyma.10G096300.3	2.5	887	907	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUCUCGACCUGAACCAAGUGA	Tradução
Glyma.10G096300.1	2.5	1222	1242	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUCUCGACCUGAACCAAGUGA	Tradução
Glyma.09G167100.1	2.5	572	592	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	CACCCGAGCUGCACCAAGCAA	Clivagem
Glyma.16G217300.1	2.5	715	735	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	CACCCGAGCUGCACCAAGCAA	Clivagem
Glyma.13G331800.1	3.0	863	883	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	CGCUCGACUUGUACGAAGCGA	Clivagem
Glyma.13G331700.1	3.0	866	886	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	CGCUCGACUUGUACGAAGCGA	Clivagem
Glyma.05G019500.1	3.0	197	217	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AACCCCAACUGCAUCAAGUGA	Clivagem
Glyma.18G276500.1	3.5	337	357	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	UUCCUCACCUGCAUCAAGUGC	Clivagem
Glyma.08G254000.1	3.5	379	399	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	UUCCUCACCUGCAUCAAGUGC	Clivagem
Glyma.10G164100.1	3.5	155	175	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	UUCCCGUUCUGCUCGAGCGA	Clivagem
Glyma.03G133700.2	3.5	4723	4743	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUUUCCACCUCACCAAGCGA	Tradução
Glyma.03G133700.1	3.5	4811	4831	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUUUCCACCUCACCAAGCGA	Tradução
Glyma.08G156800.1	3.5	4647	4667	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUUCAGAACUGCACUAAGUGA	Clivagem
Glyma.05G191600.1	3.5	4809	4829	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AUUCAGAACUGCACUAAGUGA	Clivagem
Glyma.13G133500.4	3.5	1035	1055	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	ACACUGAACUGUGCCAAGCGA	Clivagem
Glyma.13G133500.3	3.5	1189	1209	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	ACACUGAACUGUGCCAAGCGA	Clivagem
Glyma.13G133500.1	3.5	1405	1425	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	ACACUGAACUGUGCCAAGCGA	Clivagem
Glyma.13G133500.2	3.5	1472	1492	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	ACACUGAACUGUGCCAAGCGA	Clivagem
Glyma.13G287100.1	3.5	1047	1067	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	UGUUUGCCUGCAUUAAGCGA	Clivagem
Glyma.11G148100.1	3.5	270	290	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AACCCUAUUUGCACCAAACGA	Clivagem
Glyma.12G069300.1	3.5	541	561	UCGCUUGGUGCAGGUCGGGAA	:::.....:::..	AACCCUAUUUGCACCAAACGA	Clivagem
Glyma.13G337400.3	4.0	2375	2396	UCGCUUGGUGCAGGUC-GGGAA	:::.....:::..	UUCCAGGUUUGCACCAAGUGA	Clivagem
Glyma.13G337400.2	4.0	2402	2423	UCGCUUGGUGCAGGUC-GGGAA	:::.....:::..	UUCCAGGUUUGCACCAAGUGA	Clivagem
Glyma.13G337400.1	4.0	2479	2500	UCGCUUGGUGCAGGUC-GGGAA	:::.....:::..	UUCCAGGUUUGCACCAAGUGA	Clivagem
Glyma.15G037100.1	4.0	2604	2625	UCGCUUGGUGCAGGUC-GGGAA	:::.....:::..	UUCCAGGUUUGCACCAAGUGA	Clivagem

Glyma.08G077900.1	4.0	9	29	UCGCUUGGUGCAGGUCGGGAA :	UUCUCCACCUGUACCAAGUAA	Clivagem
Glyma.18G300200.2	4.0	4880	4900	UCGCUUGGUGCAGGUCGGGAA :	UUUCCGAACUGCACUUAGUGA	Clivagem
Glyma.18G300200.1	4.0	4980	5000	UCGCUUGGUGCAGGUCGGGAA :	UUUCCGAACUGCACUUAGUGA	Clivagem
Glyma.08G361500.1	4.0	4995	5015	UCGCUUGGUGCAGGUCGGGAA :	UUUCCGAACUGCACUUAGUGA	Clivagem
Glyma.08G361500.2	4.0	4995	5015	UCGCUUGGUGCAGGUCGGGAA :	UUUCCGAACUGCACUUAGUGA	Clivagem
Glyma.04G180100.5	4.0	4181	4201	UCGCUUGGUGCAGGUCGGGAA :	GUCUUGACUUGAACCAAGUGG	Tradução
Glyma.04G180100.3	4.0	4181	4201	UCGCUUGGUGCAGGUCGGGAA :	GUCUUGACUUGAACCAAGUGG	Tradução
Glyma.04G180100.4	4.0	4181	4201	UCGCUUGGUGCAGGUCGGGAA :	GUCUUGACUUGAACCAAGUGG	Tradução
Glyma.04G180100.2	4.0	4181	4201	UCGCUUGGUGCAGGUCGGGAA :	GUCUUGACUUGAACCAAGUGG	Tradução
Glyma.04G180100.1	4.0	4181	4201	UCGCUUGGUGCAGGUCGGGAA :	GUCUUGACUUGAACCAAGUGG	Tradução
Glyma.08G273100.4	4.0	1372	1392	UCGCUUGGUGCAGGUCGGGAA :	UUUGUAACCUUGAUCAAGCGA	Clivagem
Glyma.08G273100.2	4.0	1395	1415	UCGCUUGGUGCAGGUCGGGAA :	UUUGUAACCUUGAUCAAGCGA	Clivagem
Glyma.08G273100.5	4.0	1438	1458	UCGCUUGGUGCAGGUCGGGAA :	UUUGUAACCUUGAUCAAGCGA	Clivagem
Glyma.08G273100.3	4.0	1461	1481	UCGCUUGGUGCAGGUCGGGAA :	UUUGUAACCUUGAUCAAGCGA	Clivagem
Glyma.01G152400.1	4.0	90	110	UCGCUUGGUGCAGGUCGGGAA :	AUCCCGCCUUGCAUCAACUGA	Clivagem
Glyma.09G089000.2	4.0	746	766	UCGCUUGGUGCAGGUCGGGAA :	GACCCCAUCUGCACCAAGCAC	Clivagem
Glyma.09G089000.1	4.0	1123	1143	UCGCUUGGUGCAGGUCGGGAA :	GACCCCAUCUGCACCAAGCAC	Clivagem
Glyma.02G257600.3	4.0	1380	1400	UCGCUUGGUGCAGGUCGGGAA :	UUUACAAGCUGCACCAAGUGA	Clivagem
Glyma.02G257600.2	4.0	2473	2493	UCGCUUGGUGCAGGUCGGGAA :	UUUACAAGCUGCACCAAGUGA	Clivagem
Glyma.02G257600.1	4.0	2492	2512	UCGCUUGGUGCAGGUCGGGAA :	UUUACAAGCUGCACCAAGUGA	Clivagem
Glyma.02G286800.7	4.0	344	364	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.9	4.0	437	457	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.14G028100.2	4.0	440	460	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.8	4.0	344	364	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.11	4.0	505	525	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.10	4.0	437	457	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.3	4.0	739	759	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.5	4.0	726	746	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem

Glyma.02G286800.4	4.0	739	759	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.1	4.0	726	746	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.02G286800.6	4.0	895	915	UCGCUUGGUGCAGGUCGGGAA :	AGUCUGAUCUGCAUCAAGUCA	Clivagem
Glyma.08G077900.1	4.5	410	430	UCGCUUGGUGCAGGUCGGGAA	.. : .. :	ACCUGGACCUCGCCAAGUGG	Tradução
Glyma.08G005500.3	4.5	758	778	UCGCUUGGUGCAGGUCGGGAA	: .. : .. :	UACCCGACCUGCACUAAGAAU	Clivagem
Glyma.08G005500.1	4.5	758	778	UCGCUUGGUGCAGGUCGGGAA	: .. : .. :	UACCCGACCUGCACUAAGAAU	Clivagem
Glyma.08G005500.2	4.5	758	778	UCGCUUGGUGCAGGUCGGGAA	: .. : .. :	UACCCGACCUGCACUAAGAAU	Clivagem
Glyma.17G255300.1	4.5	286	306	UCGCUUGGUGCAGGUCGGGAA	.. : .. : .. :	CUCUCGACCUGUACCGAGUCC	Clivagem
Glyma.17G255300.3	4.5	286	306	UCGCUUGGUGCAGGUCGGGAA	.. : .. : .. :	CUCUCGACCUGUACCGAGUCC	Clivagem
Glyma.17G255300.2	4.5	286	306	UCGCUUGGUGCAGGUCGGGAA	.. : .. : .. :	CUCUCGACCUGUACCGAGUCC	Clivagem
Glyma.04G204600.2	4.5	480	500	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.9	4.5	585	605	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.04G204600.1	4.5	683	703	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.2	4.5	813	833	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.5	4.5	824	844	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.4	4.5	873	893	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.1	4.5	824	844	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.3	4.5	1024	1044	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.7	4.5	1068	1088	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.6	4.5	1073	1093	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.06G161000.8	4.5	1104	1124	UCGCUUGGUGCAGGUCGGGAA	: .. : .. : .. :	AACCCGACCUGCACUAAGAAU	Clivagem
Glyma.18G226500.1	4.5	7416	7436	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUCCCAAUCUGCAACAAGUGU	Clivagem
Glyma.17G102400.1	4.5	492	512	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUGCCGACCUCACCGCGCGA	Tradução
Glyma.17G110800.2	4.5	165	185	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUCUCUACUUGCGUUGAGUGA	Clivagem
Glyma.11G144600.1	4.5	1339	1359	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUCUGGAUUUGUACGAAGCGA	Clivagem
Glyma.12G038900.4	4.5	3772	3792	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.12G038900.5	4.5	3801	3821	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.12G038900.2	4.5	3772	3792	UCGCUUGGUGCAGGUCGGGAA	: : : : : : : : :	UUCCCGUUUUGUGCCAAGCCA	Clivagem

Glyma.12G038900.3	4.5	3801	3821	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.12G038900.1	4.5	3772	3792	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.04G255100.4	4.5	3649	3669	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.04G255100.6	4.5	3649	3669	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.04G255100.1	4.5	3649	3669	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.04G255100.2	4.5	4061	4081	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.04G255100.5	4.5	3992	4012	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.04G255100.3	4.5	3998	4018	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGUUUUGUGCCAAGCCA	Clivagem
Glyma.17G237800.1	4.5	536	556	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	GAACUGAUCUGCACCAAGCAG	Clivagem
Glyma.13G233000.3	4.5	3541	3561	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	GUCUUGACCUGCAGCAUGUGA	Clivagem
Glyma.13G233000.1	4.5	3741	3761	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	GUCUUGACCUGCAGCAUGUGA	Clivagem
Glyma.13G233000.2	4.5	3675	3695	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	GUCUUGACCUGCAGCAUGUGA	Clivagem
Glyma.16G071500.2	4.5	11269	11289	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	CUUUUGAUCUGCACAAAGUGG	Clivagem
Glyma.16G071500.3	4.5	12043	12063	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	CUUUUGAUCUGCACAAAGUGG	Clivagem
Glyma.16G071500.1	4.5	12043	12063	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	CUUUUGAUCUGCACAAAGUGG	Clivagem
Glyma.11G163800.1	4.5	1096	1116	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUUCAAGUUUGUACCAAGCGA	Clivagem
Glyma.17G071700.1	4.5	2442	2462	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCAAGGCUUGUACUAAGUGA	Clivagem
Glyma.04G176500.1	4.5	571	592	UCGCUUGGUGCAGG-UCGGGAA	:::~::~:	CUCCCGAGCCUGCACCAUGCGC	Clivagem
Glyma.04G176500.2	4.5	571	592	UCGCUUGGUGCAGG-UCGGGAA	:::~::~:	CUCCCGAGCCUGCACCAUGCGC	Clivagem
Glyma.17G247900.1	4.5	1076	1096	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	AUCCCGCCUGCGUGAAGCGC	Clivagem
Glyma.14G145000.1	5.0	231	253	UCGCUUGGUGCAGGU--CGGGAA	:::~::~:	UUCCCGUUAUCUGCAUCGAGCGC	Clivagem
Glyma.16G138600.1	5.0	634	654	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUUUUGACUUGGGCUAAGUGA	Tradução
Glyma.09G053300.1	5.0	115	135	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUCCCGGUCUUCACAAGGUGA	Tradução
Glyma.13G283600.1	5.0	1352	1372	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	GUCCUGACAUGCACUAAGCAU	Clivagem
Glyma.04G221000.1	5.0	31	51	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUUCCGGUCUGUGUCACGUGA	Clivagem
Glyma.02G184600.2	5.0	1768	1788	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUUCUGGUCUGCACUAGGAGG	Clivagem
Glyma.02G184600.1	5.0	2111	2131	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUUCUGGUCUGCACUAGGAGG	Clivagem
Glyma.10G099200.2	5.0	2101	2121	UCGCUUGGUGCAGGUCGGGAA	:::~::~:	UUUCUGGUCUGCACUAGGAGG	Clivagem

Glyma.10G099200.3	5.0	2126	2146	UCGCUUGGUGCAGGUCGGGAA	UUUCUGGUCUGCACUAGGAGG	Clivagem
Glyma.19G189100.5	5.0	3603	3623	UCGCUUGGUGCAGGUCGGGAA	UUCUUGAUCUGCUUCAGGUGG	Clivagem
Glyma.19G189100.2	5.0	3603	3623	UCGCUUGGUGCAGGUCGGGAA	UUCUUGAUCUGCUUCAGGUGG	Clivagem
Glyma.19G189100.3	5.0	3603	3623	UCGCUUGGUGCAGGUCGGGAA	UUCUUGAUCUGCUUCAGGUGG	Clivagem
Glyma.19G189100.1	5.0	3603	3623	UCGCUUGGUGCAGGUCGGGAA	UUCUUGAUCUGCUUCAGGUGG	Clivagem
Glyma.10G099200.1	5.0	2864	2884	UCGCUUGGUGCAGGUCGGGAA	UUUCUGGUCUGCACUAGGAGG	Clivagem
Glyma.16G025200.2	5.0	972	991	UCGCUUGGUGCAGGUCGGGAA	UUCCUGACC-GUAUUAAGCGA	Clivagem
Glyma.16G025200.1	5.0	972	991	UCGCUUGGUGCAGGUCGGGAA	UUCCUGACC-GUAUUAAGCGA	Clivagem
Glyma.16G025200.5	5.0	972	991	UCGCUUGGUGCAGGUCGGGAA	UUCCUGACC-GUAUUAAGCGA	Clivagem
Glyma.16G025200.3	5.0	972	991	UCGCUUGGUGCAGGUCGGGAA	UUCCUGACC-GUAUUAAGCGA	Clivagem
Glyma.16G025200.4	5.0	972	991	UCGCUUGGUGCAGGUCGGGAA	UUCCUGACC-GUAUUAAGCGA	Clivagem
Glyma.07G056300.1	5.0	1087	1106	UCGCUUGGUGCAGGUCGGGAA	UUCCUGACC-GUAUUAAGCGA	Clivagem
Glyma.09G089000.2	5.0	794	814	UCGCUUGGUGCAGGUCGGGAA	GACCCCAACUGCACCAGGCAU	Clivagem
Glyma.09G089000.1	5.0	1171	1191	UCGCUUGGUGCAGGUCGGGAA	GACCCCAACUGCACCAGGCAU	Clivagem
Glyma.04G085400.1	5.0	359	379	UCGCUUGGUGCAGGUCGGGAA	AUCCUGCUCUGUACCAAGCUG	Clivagem
Glyma.18G226300.1	5.0	6892	6912	UCGCUUGGUGCAGGUCGGGAA	UUCCAAUUGCAACAAGUGU	Clivagem
Glyma.19G148800.1	5.0	1186	1206	UCGCUUGGUGCAGGUCGGGAA	UUCUUGCUUGCAUAAAGUGA	Clivagem
Glyma.13G326800.2	5.0	468	488	UCGCUUGGUGCAGGUCGGGAA	UUCUGGACCUGGACCAAUGA	Tradução
Glyma.12G173500.1	5.0	462	482	UCGCUUGGUGCAGGUCGGGAA	UUCUGGACCUGGACCAAUGA	Tradução
Glyma.13G326800.1	5.0	782	802	UCGCUUGGUGCAGGUCGGGAA	UUCUGGACCUGGACCAAUGA	Tradução
Glyma.17G012200.1	5.0	47	67	UCGCUUGGUGCAGGUCGGGAA	UUCUGGGCGUGCACCAAGCAA	Clivagem
Glyma.20G120700.1	5.0	676	696	UCGCUUGGUGCAGGUCGGGAA	UUCUGGGUCUGUCCAAGUGA	Clivagem
Glyma.17G105100.1	5.0	3437	3457	UCGCUUGGUGCAGGUCGGGAA	UUGCUGACUUGCACCACCCGA	Clivagem
Glyma.17G105100.4	5.0	3500	3520	UCGCUUGGUGCAGGUCGGGAA	UUGCUGACUUGCACCACCCGA	Clivagem
Glyma.17G105100.3	5.0	3500	3520	UCGCUUGGUGCAGGUCGGGAA	UUGCUGACUUGCACCACCCGA	Clivagem
Glyma.17G105100.2	5.0	3664	3684	UCGCUUGGUGCAGGUCGGGAA	UUGCUGACUUGCACCACCCGA	Clivagem
Glyma.02G032200.4	5.0	873	893	UCGCUUGGUGCAGGUCGGGAA	CUCUCGGCCUGGAACAAGCGC	Tradução

CAPÍTULO 2

Table S1. Primer sequences (forward/reverse) used in RT-qPCR analysis.

Gene name	Sequence (5'-3')
Glyma.U027200_F	CAAGAGCATCAATAAGGGACG
Glyma.U027200_R	CAAACCGAAAGCAAACCTCC
Glyma.08G339900_F	TGTGGGCTGAAATGGAGATG
Glyma.08G339900_R	CACTGCAAGGCTGGATTTTC
AGO1_F	CAAAGCGTCACCACACAAG
AGO1_R	AGTCAAATTCGGTGGGATGG
AGO4_F	TCTCGATGAAAATTGGGAGCC
AGO4_R	TGTCGATAACAGTTCCAGGTG
DCL3_F	GCCAGTGCCTACATTTGATTC
DCL3_R	TTCTGTCCCTTGCTTCCTTG
DRM2_F	CGGTCTTTGGTGAAAATGGG
DRM2_R	CTTCAGCCTTTGCCATTTGAG
ROS1_F	GCCAACAGAGACAAATGAGC
ROS1_R	CCTAATCTGGAAAACTGCTCAG
METK4_F	CTGAGCCTTTGTCTGTGTTG
METK4_R	CCTGTTATTTCCACCCCTCTTG

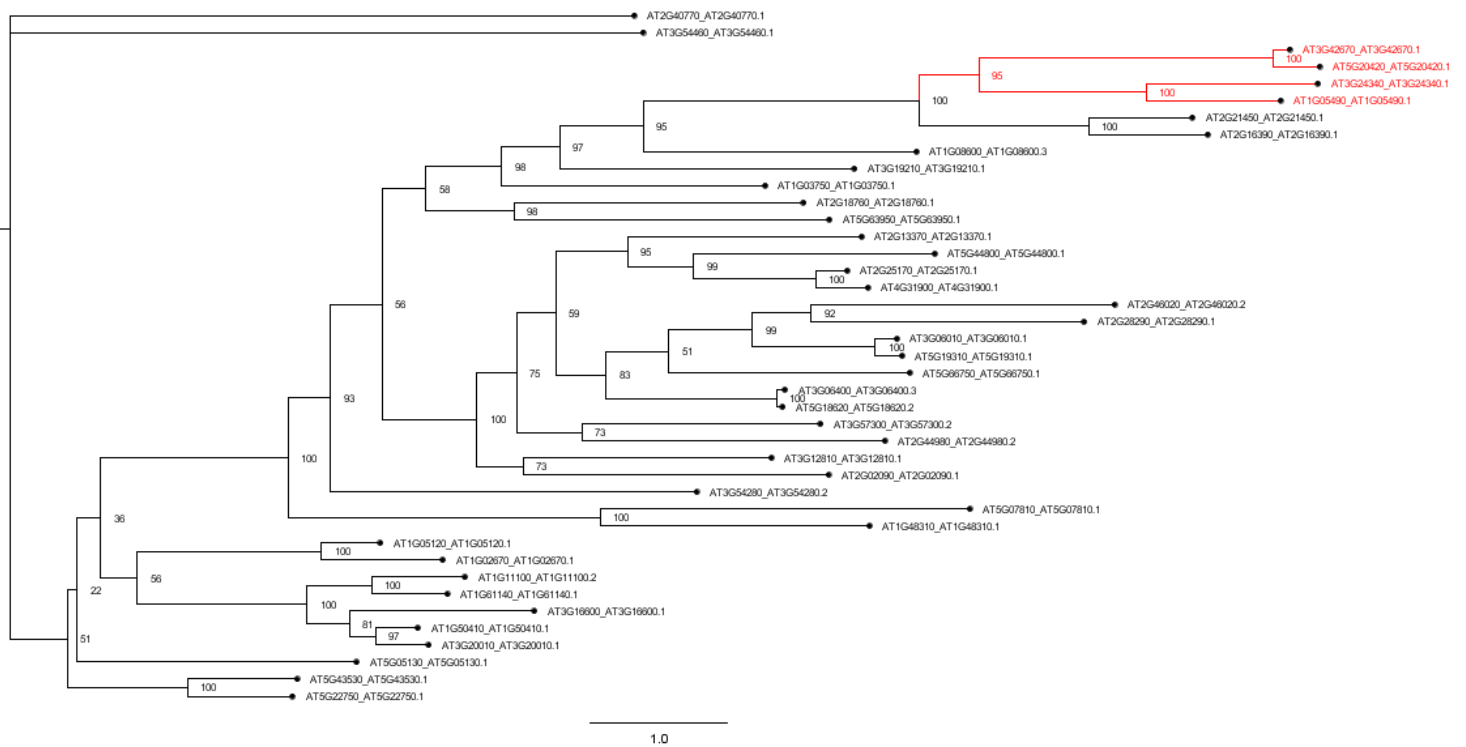


Fig. S1. Phylogenetic tree of 41 proteins containing the SNF2 domain. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.

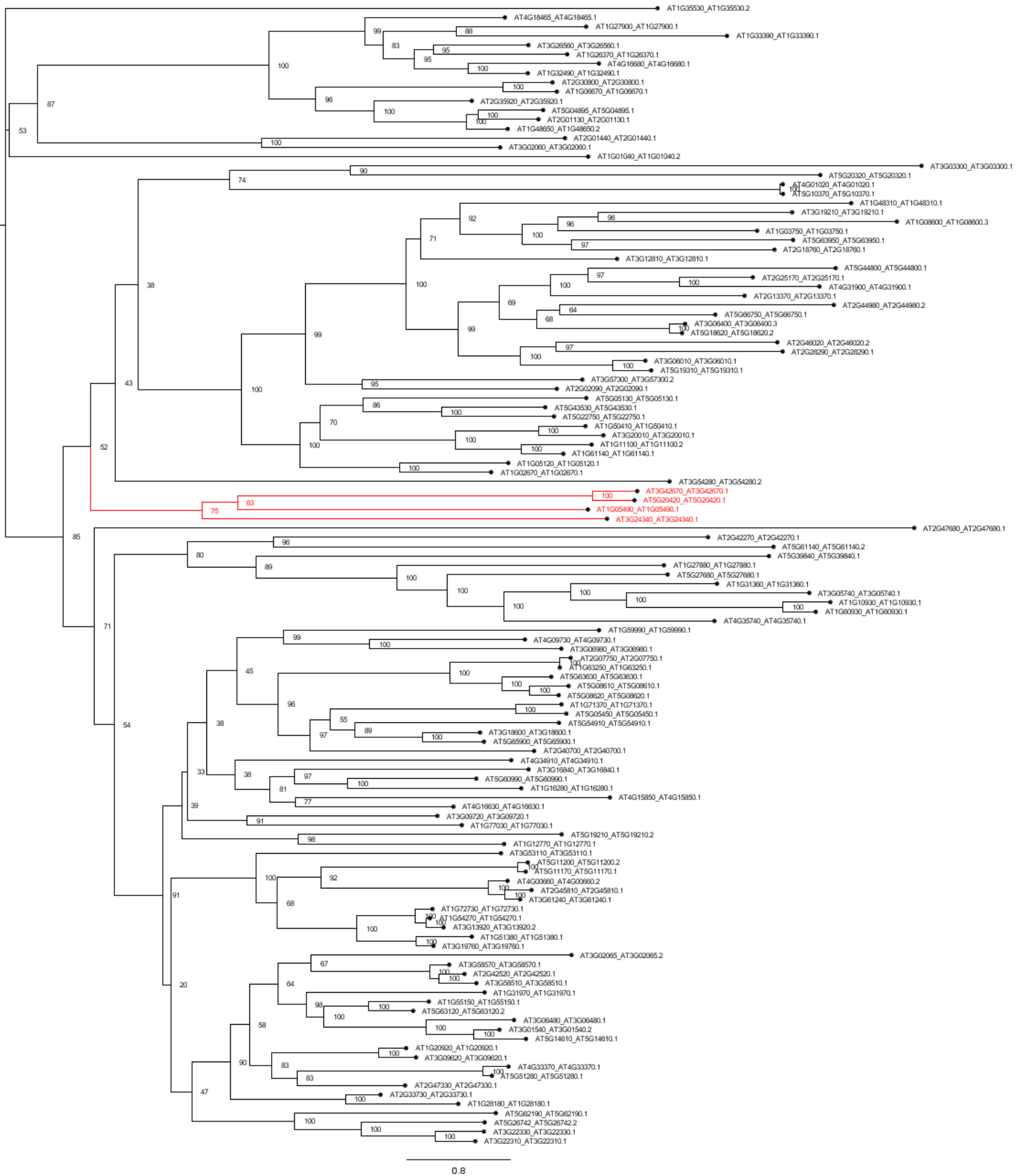


Fig. S2. Phylogenetic tree with the 123 proteins featuring the Helicase C domain. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.

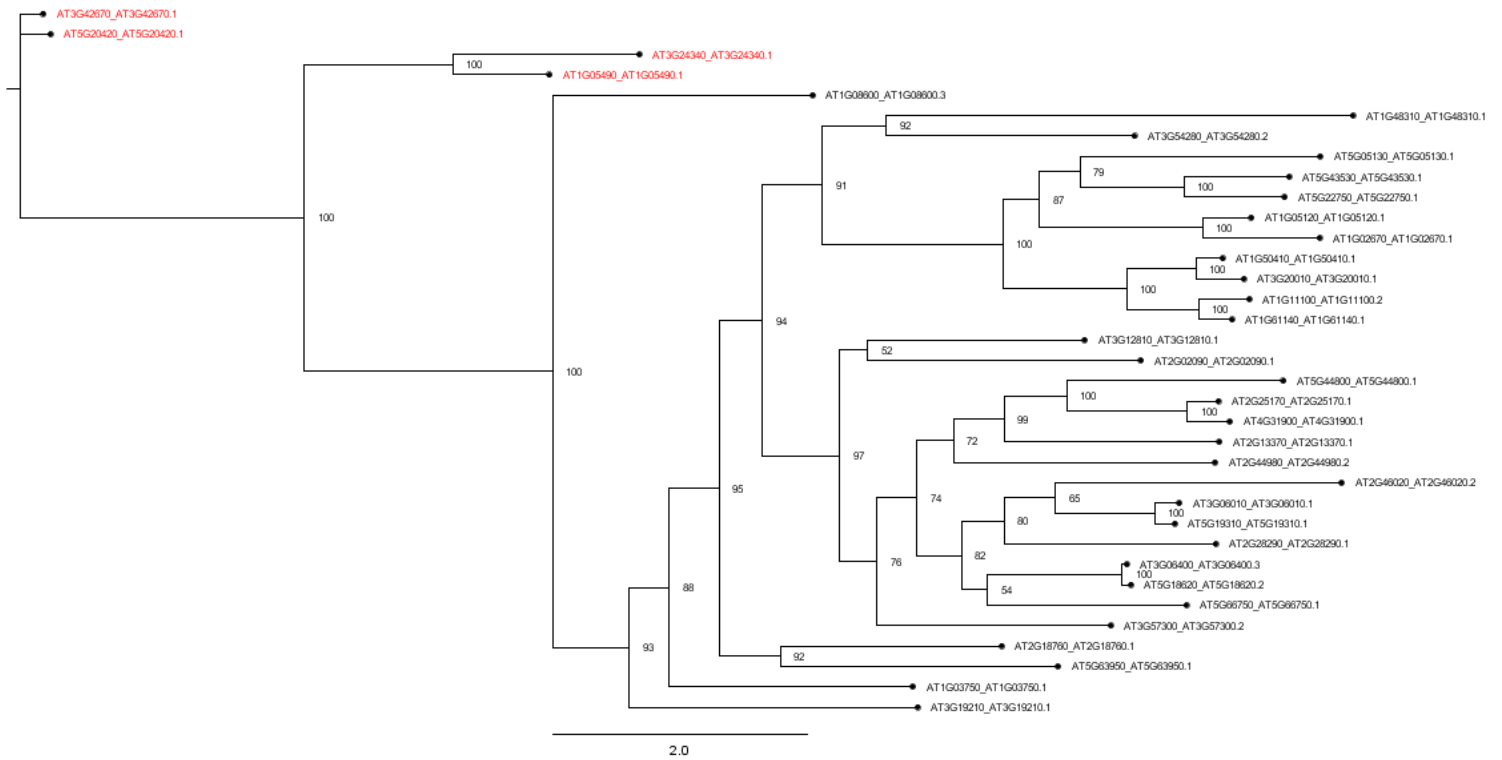


Fig. S3. Phylogenetic tree with the 35 proteins in common between the SNF2 and Helicase C domains. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.

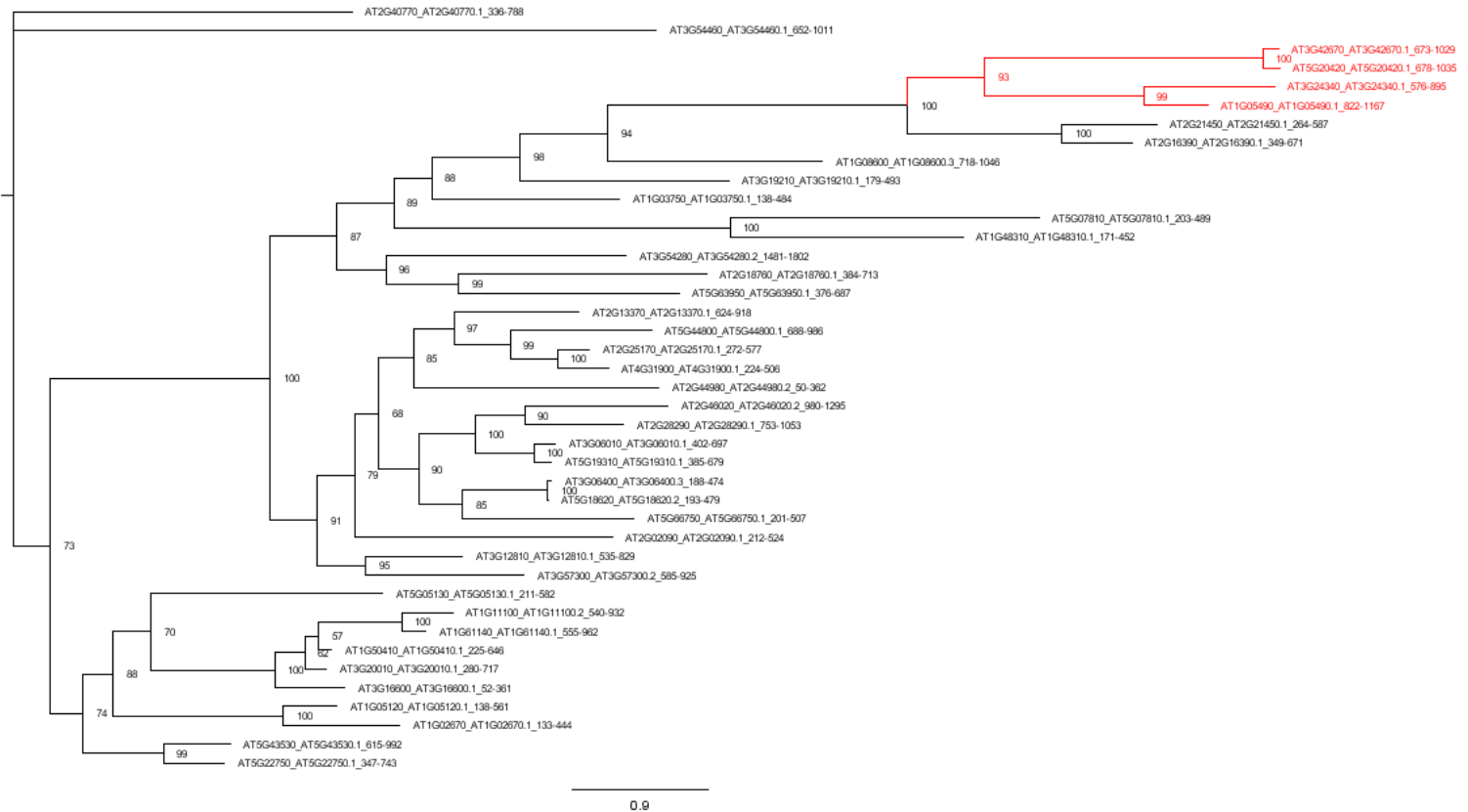


Fig. S4. Phylogenetic tree containing the 41 SNF2 domains, but excluding the Helicase C domain of the 35 proteins with both domains. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.



Fig. S5. Phylogenetic tree comprising the 123 Helicase C domains, but excluding the SNF2 domain of the 35 proteins. Arabidopsis CLSY1-4 proteins are highlighted in red. The values showed on the branches represent the bootstrap values from 1000 replicates.

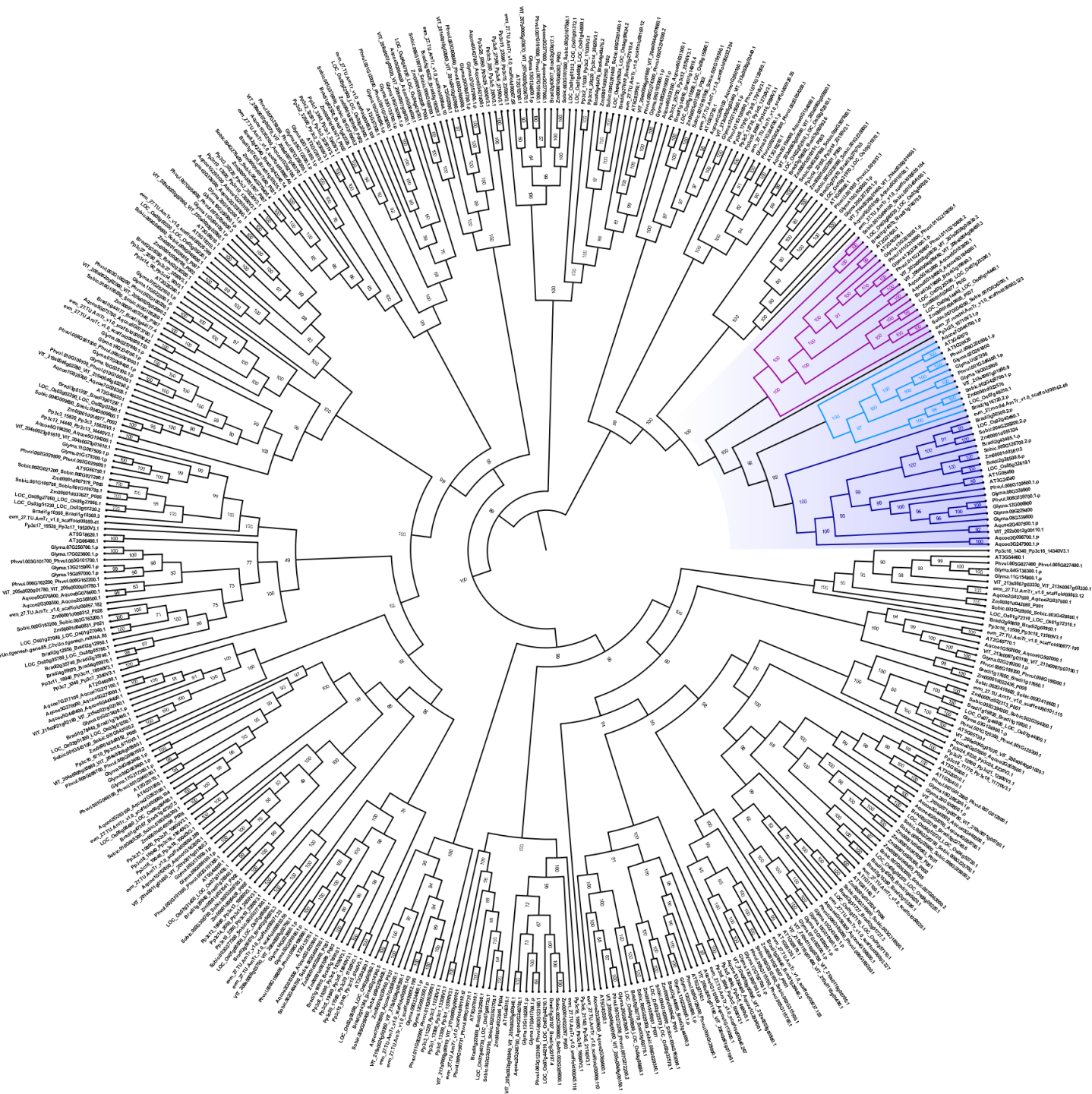


Fig. S6. Complete phylogenetic tree of the 447 identified proteins, with emphasis on the monophyletic group containing proteins related to the CLSY and DRD1 families. Clades 1, 2, and 3 are shown in light blue, dark blue, and purple, respectively. The values indicated on the branches represent the bootstrap values from 1000 replicates.

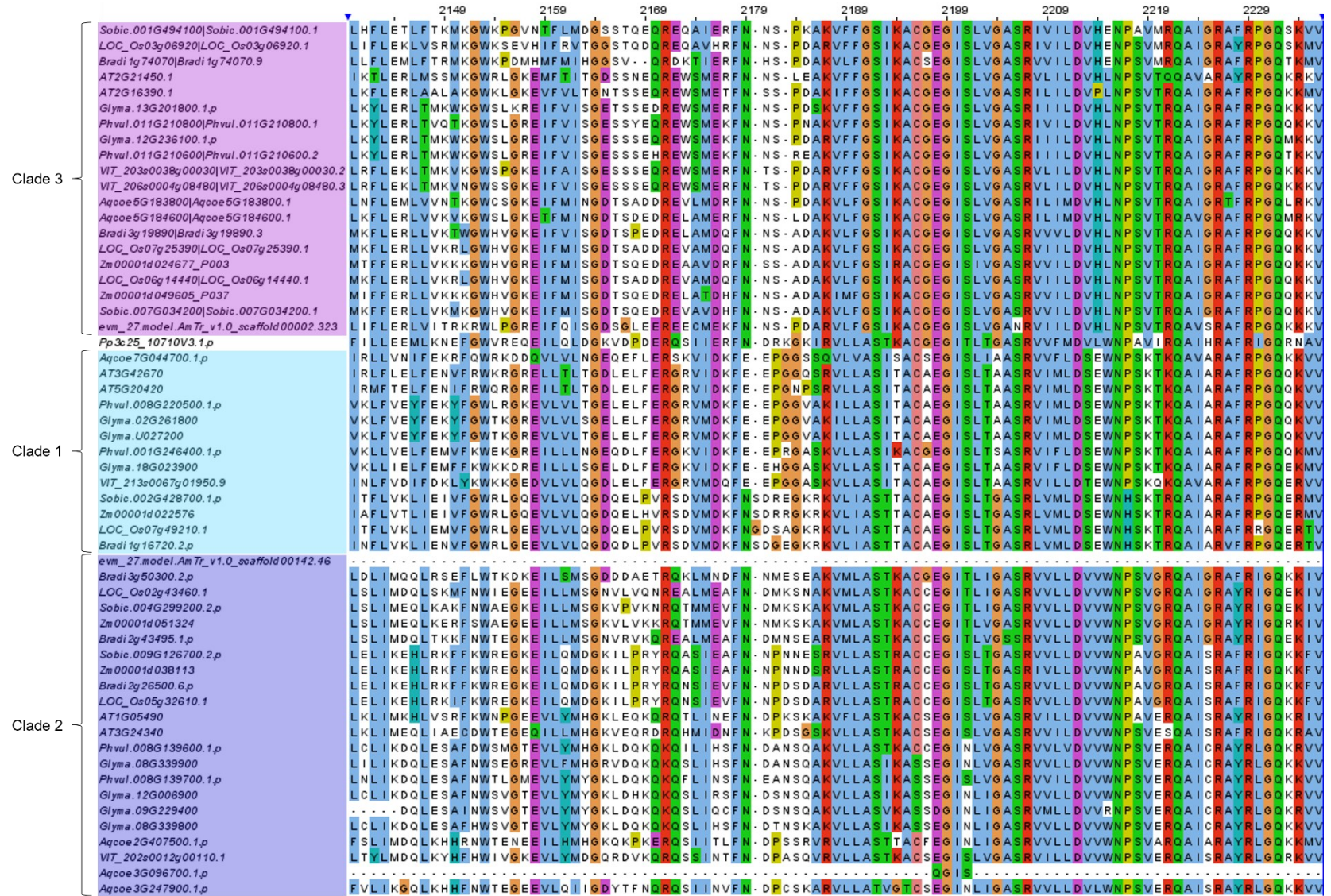


Fig. S7. Multiple sequence alignment of the Helicase C domain obtained in Jalview, highlighting the clades 1, 2, and 3. The clades are in the order in which they were arranged in the phylogenetic tree (Fig. S6). A color scale on amino acids was used to indicate sequence similarity.

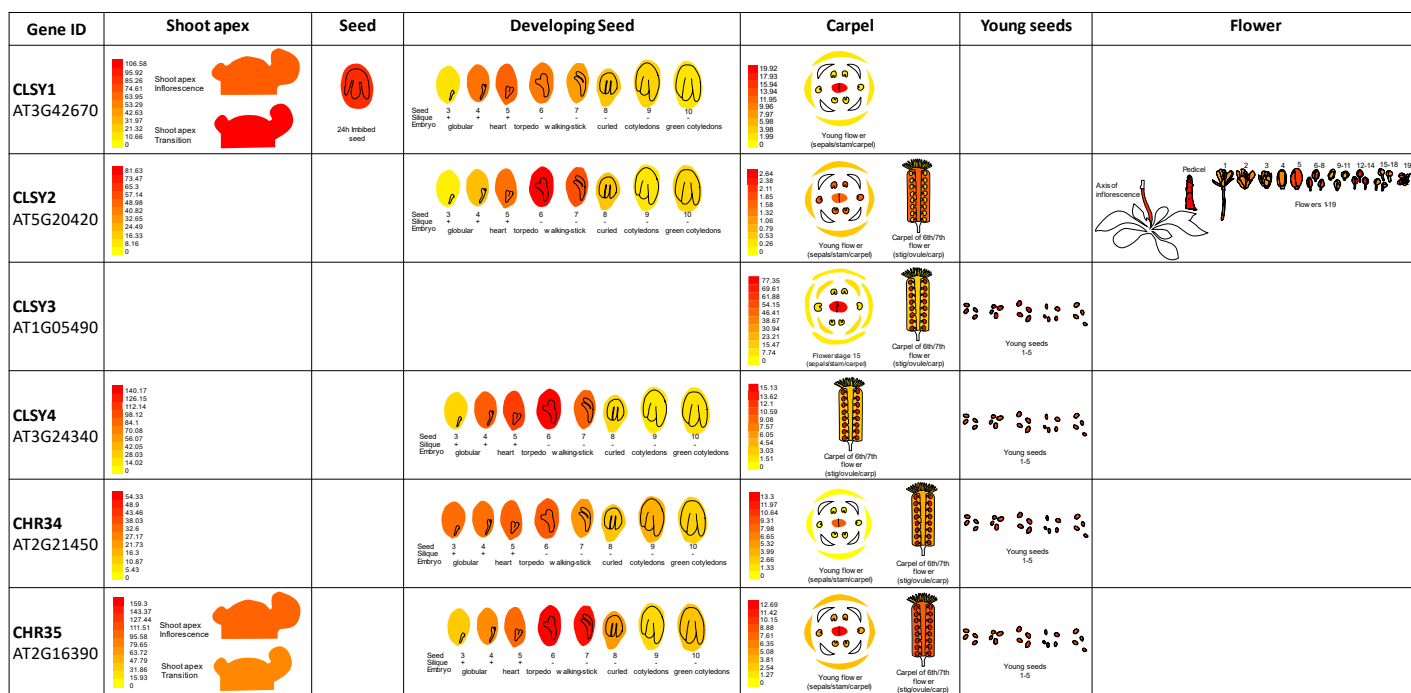


Fig. S8. Tissues in which the CLSY1-4, AT2G16390/DRD1, and AT2G21450 genes are most expressed in Arabidopsis. The scale in the Shoot Apex column refers to the expression of genes in shoot apex, seed, and developing seed tissues. The scale in the carpel column refers to the expression of genes in carpel, young seeds, and flower tissues.

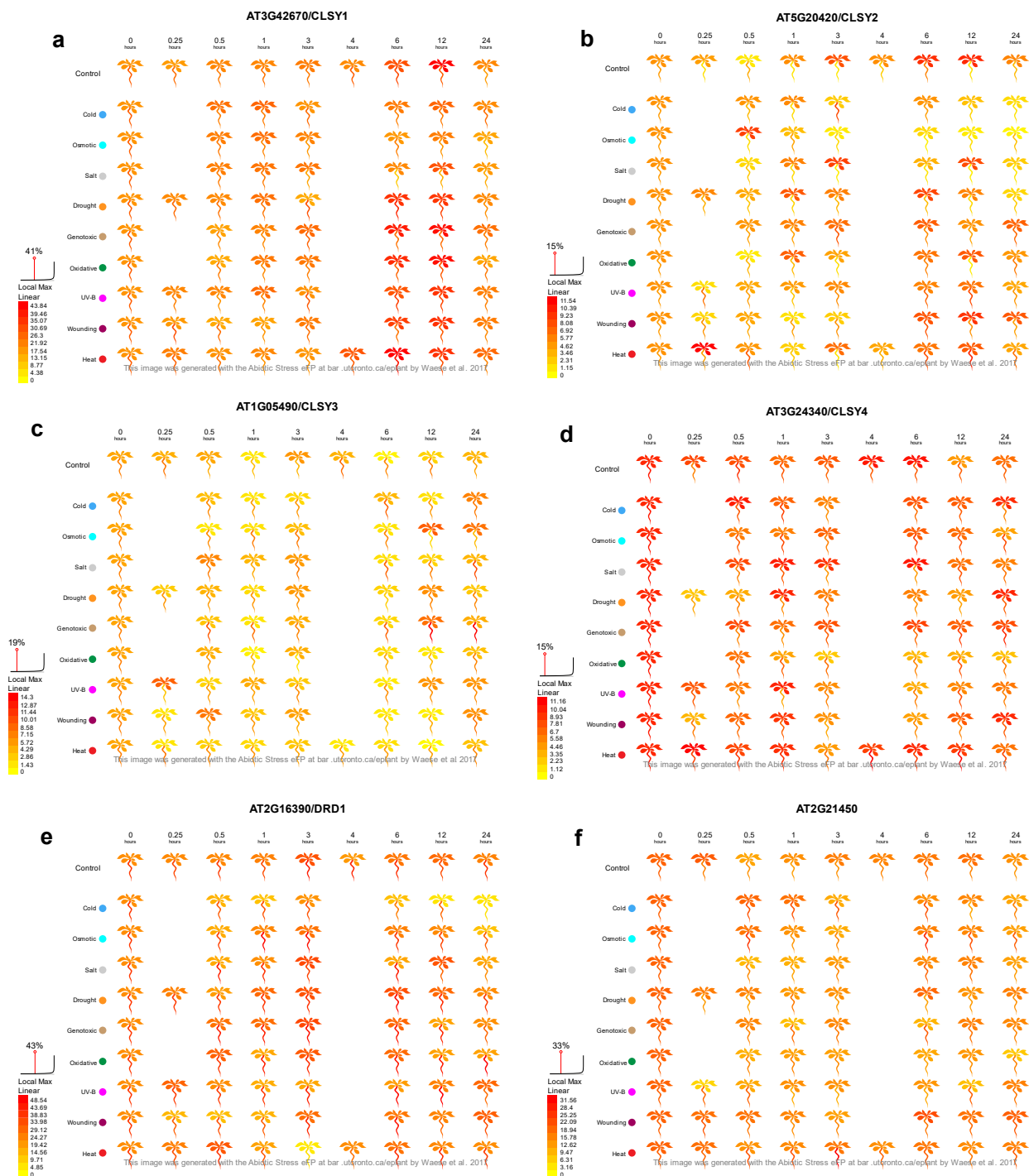


Fig. S9. Expression profile of Arabidopsis genes CLSY1-4, AT2G16390/DRD1, and AT2G21450 under nine types of abiotic stresses in shoot and root tissues. (a) CLSY1; (b) CLSY2; (c) CLSY3; (d) CLSY4; (e) AT2G16390/DRD1; (f) AT2G21450.

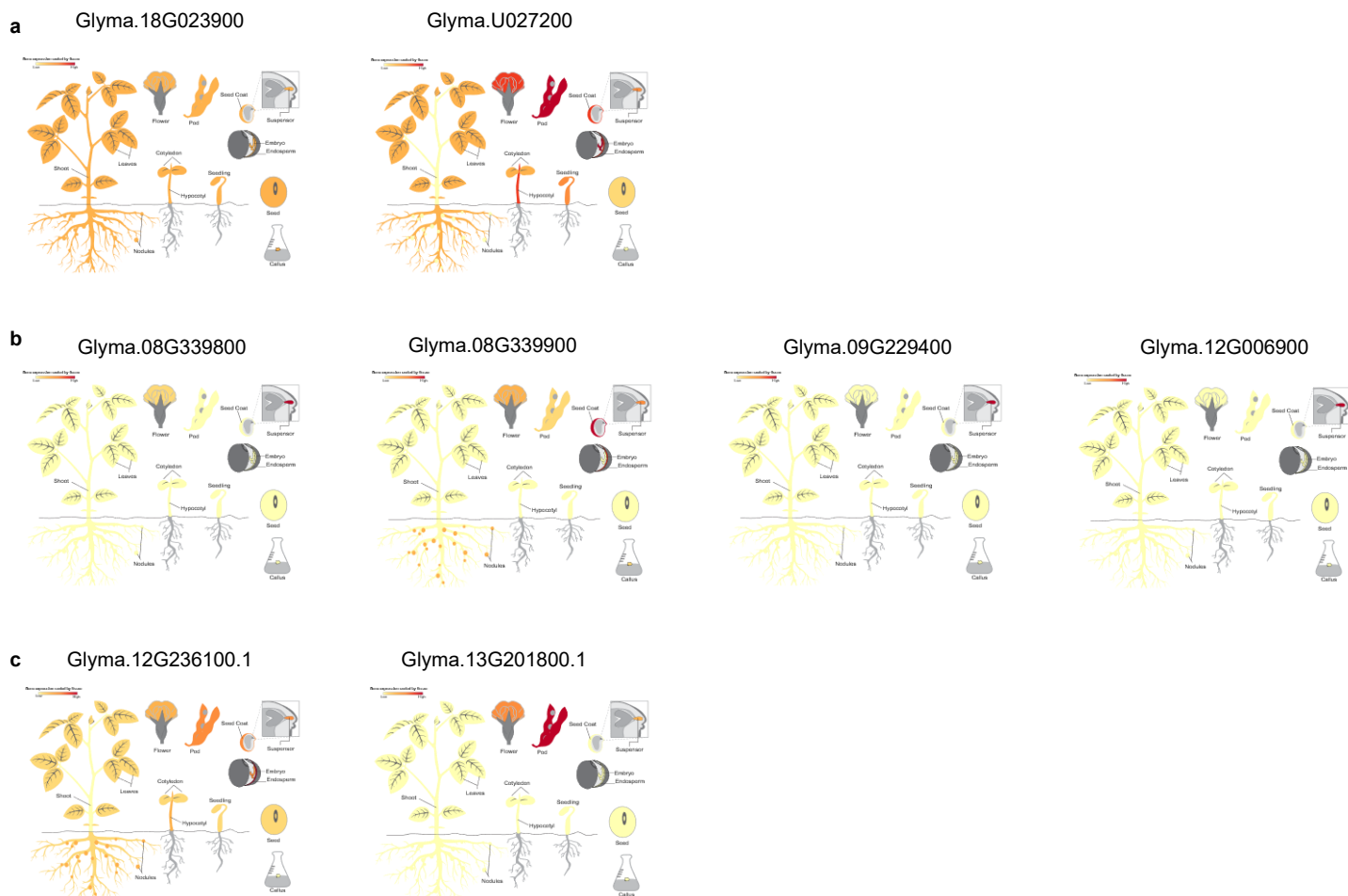


Fig. S10. Expression profiles of soybean genes from the three clades in 14 different tissues. (a) Clade 1 genes; (b) Clade 2 genes; (c) Clade 3 genes. No results were reported for the gene Glyma.02G261800.

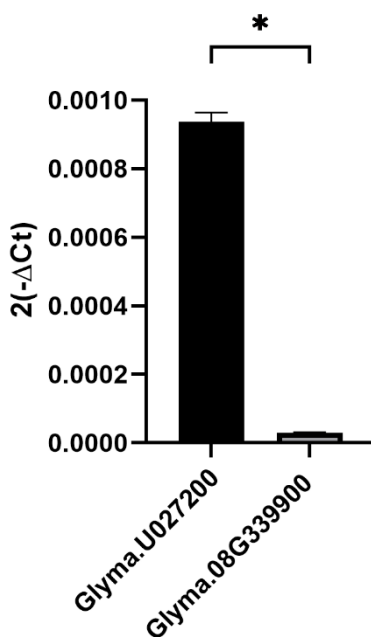


Fig. S11. Relative expression of Glyma.U027200 and Glyma.08G339900 genes in embryonic axes of cultivar BRS284, which refers to the same transcriptome used in the Soybean Expression Atlas database. The asterisk indicates significantly different expression values between genes according to the t-test ($P < 0.05$).

ANEXO

Artigo produzido durante o doutorado, à parte da tese, publicado na revista Computational Biology and Chemistry (2022).

Research article

***In silico* identification of candidate miRNA-encoded Peptides in four Fabaceae species**

Paula Machado de Araújo¹, Clícia Grativol^{1*}

¹Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, Brazil.

*Corresponding author.

<https://doi.org/10.1016/j.compbiolchem.2022.107644>

Abstract

MicroRNAs (miRNAs) are one of the main regulators of gene expression. Recent studies have demonstrated that primary transcripts of miRNAs (pri-miRNAs) encode regulatory peptides, called miRNA-encoded peptides (miPEPs), capable of enhancing the expression of their associated miRNAs in plants. In this work, we aimed to computationally identify miPEPs produced by small open reading frames (ORFs) in pri-miRNAs from four species of Fabaceae. Five families of miRNAs were investigated, based on their role in plant-microorganism interaction. We used the miR171 family as a training dataset centered on the information about mtr-miPEP171b and vvi-miPEP171d already described. From the sequences of the pri-miRNAs and the genomic regions where they were located, ORFs encoding putative miPEPs were predicted. The 5'-most ORFs encoding peptides on pri-miRNAs were aligned and the amino acids conservation was observed. In total, 81 sequences of potential miPEPs were identified. We found conserved miPEPs inside pri-miRNAs from soybean and between soybean, common bean, and cowpea. Besides, our results showed few conserved miPEPs among isoforms of the same miRNA and no conservation among different miRNA families, which indicate the possible specificity of miPEPs in relation to their corresponding miRNAs. Our findings contribute to the understanding of miPEPs features in plants and provide the basis for studies aiming the biotechnological use of miPEPs in leguminous species.

Keywords: MicroRNAs; miPEPs; *in silico* analysis; Fabaceae.