

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE BIOCÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

**RAHISA HELENA DA SILVA**

**GENÔMICA ESTRUTURAL E FUNCIONAL DE GENES CORREGULADORES DE  
CANA-DE-AÇÚCAR EM RESPOSTA À SUPRESSÃO DE REGA**

Recife

2020

**RAHISA HELENA DA SILVA**

**GENÔMICA ESTRUTURAL E FUNCIONAL DE GENES CORREGULADORES DE  
CANA-DE-AÇÚCAR EM RESPOSTA À SUPRESSÃO DE REGA**

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

**Orientador:** Prof°. Dr. Éderson Akio Kido

Recife

2020

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

Silva, Rahisa Helena da  
Genômica estrutural e funcional de genes correguladores de cana-de-açúcar em resposta a supressão de rega / Rahisa Helena da Silva. - 2020.

100 f. : il.

Orientador: Prof. Dr. Éderson Akio Kido.  
Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Genética, Recife, 2020. Inclui referências e anexos.

1. Cana-de-açúcar. 2. Cana-de-açúcar – Melhoramento genético. 3. Genética vegetal. I. Kido, Éderson Akio (orientador). II. Título.

633.61

CDD (22.ed.)

UFPE/CB-2020-088

**RAHISA HELENA DA SILVA**

**GENÔMICA ESTRUTURAL E FUNCIONAL DE GENES CORREGULADORES DE  
CANA-DE-AÇÚCAR EM RESPOSTA À SUPRESSÃO DE REGA**

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

**Aprovado em: 20/02/20**

**BANCA EXAMINADORA:**

---

Profº. Dr. Éderson Akio Kido (orientador)

Universidade Federal de Pernambuco

---

Profº. Dr. Valdir Queiroz Balbino (Examinador Interno)

Universidade Federal de Pernambuco

---

Profª. Drª. Flávia Figueira Aburjaile (Examinador Externo)

Universidade Federal de Pernambuco

---

Dr. José Ribamar Costa Ferreira Neto (Examinador Externo)

Universidade Federal de Pernambuco

## **AGRADECIMENTOS**

Agradeço ao **Dr. Prof. Éderson Akio Kido**, por me iniciar na ciência e ter me orientado até o momento da entrega desta dissertação.

Agradeço ao suporte financeiro oferecido pelas agências de fomento, em especial ao **Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)**.

Agradeço a meus pais, por terem dado apoio psicológico, e motivação necessários para conclusão desta etapa acadêmica.

Agradeço ao meu esposo, **Manassés Daniel da Silva**, pelo apoio, paciência, amor e carinho dados a mim durante esse período.

Agradeço a todos os membros do Laboratório de Genética Molecular, em especial **Bruna, Elvia, Francielly e Vanessa**

## RESUMO

A cana-de-açúcar (*Saccharum* spp.) é uma cultura de elevada importância econômica, porém, sua produtividade é fortemente ameaçada pela seca. Plantas submetidas a estresses modulam seu transcriptoma com o intuito de minimizar os danos decorrentes. Os genes que atuam na regulação da expressão gênica atuam de frente a esta resposta, dentre estes, os correguladores (CoREGs) carecem de estudos acerca de como se comportam em plantas sob estresse. O presente trabalho representa o primeiro CoREGoma descrito para cana-de-açúcar, que abrange todas as suas 24 famílias e composto por 3.800 transcritos. Uma metodologia de transcriptômica (HT-SuperSAGE) foi aplicada a acessos de cana-de-açúcar contrastantes (tolerante e sensível) submetidos a supressão de rega (24 h) e permitiu levantar o perfil transcricional dos CoREGs nas raízes dos acessos frente ao estímulo aplicado. Os CoREGs identificados foram submetidos a análises de enriquecimento de termos de ontologia gênica (GO) e de fatores de transcrição. Os termos GO confirmaram a natureza indireta da atuação do grupo na regulação da expressão gênica, bem como, fatores de transcrição associados a respostas a estresses abióticos foram preditos a regular CoREGs associados as *unitags* diferencialmente expressas. Prováveis CoREGs ortólogos foram distribuídos em todos os cromossomos de sorgo, milho e arroz. A comparação da distribuição dos *loci* por cromossomos tendo sorgo por referência mostrou um maior grau de microssintenia entre sorgo e arroz, mesmo sendo arroz mais distante evolutivamente quando comparado com milho. Os CoREGs de expressões validadas por qPCR abrangem vias importantes frente a estresses abióticos e muito provavelmente contribuem para a melhor performance dos acessos tolerantes frente ao estímulo aplicado.

Palavras-chave: Transcriptômica. Regulação da expressão gênica. Estresse abiótico.

## ABSTRACT

Sugarcane (*Saccharum* spp.) is a crop of high economic importance, however, its productivity is strongly threatened by drought. Plants subjected to stress modulate their transcriptome in order to minimize the resulting damage. The genes that act in the regulation of gene expression act against this response, among them, the co-regulators (CoREGs) lack studies on how they behave in plants under stress. The present work represents the first CoREGoma described for sugarcane, which covers all 24 families and consists of 3,800 transcripts. A transcriptomic methodology (HT-SuperSAGE) was applied to contrasting sugarcane accessions (tolerant and sensitive) submitted to irrigation suppression (24 h) and allowed to raise the transcriptional profile of CoREGs at the roots of the accessions in view of the applied stimulus. The identified CoREGs were submitted to enrichment analyzes of gene ontology (GO) terms and transcription factors. The terms GO confirmed the indirect nature of the group's role in regulating gene expression, as well as, transcription factors associated with responses to abiotic stresses were predicted to regulate CoREGs associated with differentially expressed unitags. Probable orthologous CoREGs were distributed on all sorghum, maize and rice chromosomes. Comparison of loci distribution by chromosomes, with sorghum being the reference, showed a higher degree of microsynteny between sorghum and rice, even though rice is evolutionarily more distant when compared to maize. The qPCR-validated CoREGs cover important pathways against abiotic stresses and most likely contribute to the better performance of tolerant accesses against the applied stimulus.

Keywords: Transcriptomics. Regulation of gene expression. Abiotic Stress.

## LISTA DE FIGURAS

Figura 1 -	Média da produção (1994 – 2017) das culturas mais produtivas (milhões de toneladas) .....	23
Figura 2 -	Média da produção (1994 – 2017) da cana-de-açúcar entre os 10 países mais produtores.....	23
Figura 3 -	DNA compactado em nucleossomos.....	30
Figura 4 -	Esquema das modificações pós-traducionais nas caudas de histonas. Para fins de simplificação estão ilustrados apenas sítios de acetilação, metilação, fosforilação e ubiquitinação.....	31
Figura 5 -	Representação esquemática da ação de HATs e HDACs na regulação de respostas a estresses abióticos.....	34
Figura 6 -	Representação esquemática do processo adição (HMTs), remoção (HDMTs) e interpretação de grupos metil adicionados a caudas de histonas.....	35
Figura 7 -	Os diferentes resultados da ação de complexos de remodelamento da cromatina (CRCs).....	39
Figura 8 -	Representação esquemática sintetizada do Dogma Central da Biologia Molecular.....	41
Figura 9 -	Fluxograma da técnica HT-SuperSAGE.....	44
Figura 10 -	Distribuição das ESTs de cana-de-açúcar nas famílias de corretores oriundas do iTAK ( <a href="http://itak.feilab.net/cgi-bin/itak/index.cgi">http://itak.feilab.net/cgi-bin/itak/index.cgi</a> ).....	56
Figura 11 -	Diagramas de Venn ilustrando <i>unitags</i> exclusivas e compartilhadas entre os dois <i>bulks</i> de acessos (SD24T vs C[TC] e SD24S vs C[SC]) entre aquelas diferencialmente expressas (DEs $p\text{-value} < 0,05$ ).....	58
Figura 12 -	Distribuição das <i>unitags</i> HT-SuperSAGE nas famílias CoREG identificadas nas ESTs de cana-de-açúcar.....	59

- Figura 13 - Clusterização hierárquica feita a partir dos dados de modulação da expressão (FC; *Fold Change*) nos dois *bulks* de cultivares, de parte das *unitags* DE (diferencialmente expressas,  $p\text{-value} < 0,0001$ )..... 62
- Figura 14 - Diagramas de Venn da distribuição de termos GO enriquecidos de alvos associados a *unitags* DE nos contrastes tolerante (SD24TvsC [TC]) e sensível (SD24SvsC [SC])..... 64
- Figura 15 - Famílias de FTs super-representados na análise de enriquecimento ( $p\text{-value threshold} \leq 0,01$ ), considerando as *unitags* DE ( $p < 0,05$ ) da resposta tolerante (a) e sensível (b), de cana-de-açúcar (quatro acessos cada), após 24h de supressão de rega..... 67
- Figura 16 - Quantidade de *loci* associados a corre reguladores em *S. bicolor*, *Z. mays* e *O. sativa* (a). Diagrama de Venn ilustrando o compartilhamento entre ESTs de cana-de-açúcar associadas a *loci* corre reguladores entre *S. bicolor*, *Z. mays* e *O. sativa* (b)..... 70
- Figura 17 - Ideogramas circulares construídos a partir dos dados de localização cromossômica de genes corre reguladores ortólogos entre sorgo e cana-de-açúcar nas comparações (via BLASTp  $e\text{-value} e^{-10}$ ) sorgo com milho (a) e sorgo com arroz (b)..... 73

## LISTA DE TABELAS

Tabela 1 -	Alguns exemplos de bancos de dados utilizados em transcriptômica vegetal.....	46
Tabela 2 -	Total de <i>unitags</i> HT-SuperSAGE diferencialmente expressas (DE; <i>p-value</i> < 0.05), associadas a ESTs CoREGs, em raízes de cana-de-açúcar após 24 h de supressão de rega, quando comparadas a expressão dos respectivos controles irrigados.....	58
Tabela 3 -	Número de <i>unitags</i> associadas a famílias de corre reguladores em biblioteca HT-SuperSAGE de raízes de acessos contrastantes (tolerante [SD24T vs SDTC]; sensível [SD24S vs SDSC]) de cana-de-açúcar após 24 h de supressão rega. As <i>unitags</i> são representadas com base na regulação detectada.....	61
Tabela 4 -	Termos GO enriquecidos ( $p < 0,01$ ) associados a <i>unitags</i> DE ( $p < 0,05$ ) do contraste tolerante (TC[SD24T vs C]).....	66
Tabela 5-	Termos GO enriquecidos ( $p < 0,01$ ) associados a <i>unitags</i> DE ( $p < 0,05$ ) do contraste sensível (SC[SD24S vs C]).....	67
Tabela 6 -	Fatores de transcrição super-representados e seu número de regulações preditas para potenciais genes de corre reguladores associados a <i>unitags</i> DE ( <i>p-value</i> < 0,05) da resposta tolerante (SD24T vs C) e sensível (SD24C vs C).....	69
Tabela 7 -	Número de <i>loci</i> relacionados a CoREGs por espécie (sorgo, milho e arroz) e por família CoREG identificada.....	72
Tabela 8 -	Dados de expressão <i>in silico</i> dos candidatos selecionados de <i>unitags</i> HT-SuperSAGE de cana-de-açúcar expressas pelos <i>bulks</i> de cultivares contrastantes (tolerante e sensível) e respectivas expressões obtidas via RT-qPCR, em raízes após estímulo (24 h de supressão de rega).....	76

## LISTA DE ABREVIATURA E SIGLAS

<sup>1</sup> O <sub>2</sub>	Oxigênio singleto
ABA	Ácido abscísico
ABRE	<i>ABA responsive element</i> (Elemento ABA responsivo)
ADA2b	<i>Transcriptional adapter</i> (Adaptador transcricional ADA2b)
ADMA	<i>Asymmetric dimethyl groups</i> (Grupos di-metil assimétricos)
AREB/ABF	<i>ABRE binding factors</i> (Fatores ligantes de ABRE)
ARID	<i>AT-rich Interactive Domain</i> (Domínio Interativo Rico em AT)
ATX1	<i>Arabidopsis Trithorax-like</i>
BBH	<i>Bidirecional Best-Hit</i> (Best-Hit Bidirecional)
BER	<i>Base Excision Repair</i> (Reparo por Excisão de Base)
BLAST	<i>Basic Local Alignment Search Tool</i> (Ferramenta Básica de Alinhamento Local)
CA	Anidrase carbônica
cDNA	complementary DNA (DNA complementar)
CHD	Chromo Helicase Domain (Helicase contendo Domínio Cromo)
CIA	<i>Chloroplast Import Apparatus</i> (Aparato de Importação do Cloroplasto)
CO <sub>2</sub>	Dióxido de carbono
CONAB	Companhia Nacional de Abastecimento
CoREGs	Correguladores
Co-REST	<i>Chromatin transcriptional repressor associated</i> (Repressor Transcricional Associado a Cromatina)
CRC	Complexo de Remodelamento da Cromatina

CTC	Centro de Tecnologia Canavieira
CTR1	<i>Constitutive Triple Response 1</i> (Proteína de Resposta Tripla Constitutiva)
DDBJ	DNA <i>Data Bank of Japan</i> (Banco de Dados do Japão)
DE	Diferencialmente Expresso
DNA	Ácido desoxirribonucléico
DOT1	<i>Disruptor of Telomeric Silencing 1</i> (Disruptor de Silenciamento Telomérico 1)
DR	Down regulated (Reprimido)
DRE/CRT	<i>Drought Responsive Element/ C-repeat</i> (Elemento Responsivo a Desidratação Contendo Repetição de Citosinas)
DREB1	<i>Drought Responsive Element Binding Protein 1</i> (Proteína 1 de Ligação ao Elemento Responsivo à Desidratação)
DREB2	<i>Drought Responsive Element Binding Protein 2</i> (Proteína 2 de Ligação ao Elemento Responsivo à Desidratação)
EST	<i>Expressed Sequence Tag</i> (Etiqueta de Sequência Expressa)
FAD	<i>Flavin Adenine Dependent Dinucleotide</i> (Flavina Adenina Dinucleotídeo Dependente)
FAO	<i>Food and Agriculture Organization of the United Nations</i> (Organização das Nações Unidas para Alimento e Agricultura)
GNAT/GCN5	<i>Non-repressible General Control-5 related to N-terminal Acetyltransferase</i> (Controle Geral não Repressível-5 relacionado à Acetiltransferase N-terminal)
GO	<i>Gene Ontology</i> (Ontologia Gênica)
Grassius	<i>Grass Regulatory Information Server</i>

GRN	<i>Gene Regulatory Networks</i> (Redes de Regulação Gênica)
GSNO	S-nitrosoglutathione
H2D	<i>Histone Deacetylase 2</i> (Histona desacetilase 2)
HAT	<i>Histone Acetyltransferase</i> (Histona acetiltransferase)
HDAC	<i>Histone Deacetylases</i> (Histona desacetilase)
HKMT	<i>Histone Lysine Methyltransferases</i> (Histona Lisina Metiltransferase)
HMD	<i>Histone Demethylases</i> (Histona demetilase)
HMG	<i>High Mobility Group</i> (Grupo de alta mobilidade)
HMM	<i>Hidden Markov Model</i> (Modelo oculto de Markov)
HMT	<i>Histone methyltransferases</i> (Histona metiltransferase)
HSP	<i>High Score Pair</i> (Par de alto score)
HSP101	<i>Heat Shock Protein 101</i> (Proteína de choque térmico 101)
HT-SuperSAGE	<i>High-Throughput-SuperSAGE</i> (SuperSAGE de alto rendimento)
IAC	Instituto Agronômico de Campinas
INO80/SWR	<i>Inositol Auxotroph 80</i>
INSD	<i>International Nucleotide Sequence Database</i> (Banco Internacional de Sequências Nucleotídicas)
ISWI	<i>Imitation Switch</i>
iTAK	<i>Plant Transcription factor &amp; Protein Kinase Identifier and Classifier</i>
JHDM1A	<i>Jumonji Histone Demethylase 1A</i> (Histona Demetilase 1 <sup>a</sup> contendo o Domínio JmjC)
JmjC	Jumonji
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>

LongSAGE	<i>Long Serial Analysis of Gene Expression</i>
LSD1/KDM1	<i>Demethylase 1 Lysine-specific</i> (Demetilase 1 específica de lisina)
LSH	<i>Lymphocyte-specific Helicase</i> (Helicase Específica do Linfócito)
LSM4	<i>Sm-like protein 4</i> (Proteína 4 semelhante a Sm)
MACE	<i>Massive analysis of cDNA ends</i>
MAL	Malato
MBF1	<i>Multi-Bridging Factor 1</i>
MDH	Malato desidrogenase
MET1	<i>Methyltransferase 1</i> (Metiltransferase 1)
mTERF	<i>Mitochondrial Transcription Termination Factor</i> (Fator de Terminação da Transcrição Mitocondrial)
MYST	MOZ, Ybf2/Sas3, Sas2, e Tip60
N2O3	Trióxido de dinitrogênio
NCBI	<i>National Center for Biotechnology Information</i> (Centro Nacional para Informações sobre Biotecnologia)
NGS	<i>New Generation Sequencing</i> (Sequenciamento de Nova Geração)
NO	Óxido nítrico
NO2	Dióxido de nitrogênio
npcRNAs	<i>non-protein-coding RNAs</i> (RNAs não codificantes de proteínas)
O2°-	Radical superóxido
OAA	Ácido oxaloacetato
OH°	Radical hidróxido
ONOO°	Peroxinitrito
p300/CREB	<i>cAMP Responsive Binding Protein</i> (Proteína de ligação ao

	elemento c-AMP responsivo)
PEP <sub>c</sub>	Fosfoenolpiruvato carboxilase
PHD	<i>Plant homeodomain</i>
PRMT	<i>Arginine Methyltransferase Proteins</i> (Metiltransferase de Arginina)
PSI-BLAST	<i>Position-Specific Iterated BLAST</i>
RCD1-Like	<i>Radical-induced Cell Death1</i> (Radical 1 de indução de morte celular)
RNS	<i>Reactive Nitrogen Species</i> (Espécies Reativas de Nitrogênio)
ROS	<i>Reactive Oxygen Species</i> (Espécies Reativas de Oxigênio)
RPD3/HDA1	<i>Protein-3 of Reduced post-clearance / Histone deacetylase-1</i> (Proteína-3 de dependência reduzida de postássio/histona desacetilase-1)
RSNOs	S-nitrosotióis
SAGE	<i>Serial Analysis of Gene Expression</i>
SET	<i>Su[<i>var</i>], Enhancer of Zeste, Trithorax</i>
SIR2	<i>Silence Information Regulation 2</i> (Proteína 2 Reguladora do Silenciamento de Informação)
SUCEST	<i>Sugarcane EST Project</i>
SuperSAGE	<i>Super Serial Analysis of Gene Expression</i>
TAFII250	<i>TATA Association Factor</i> (Fator de associação à TATA)
TAIR	<i>The Arabidopsis Information Resource</i> (Recurso de informação para Arabidopsis)
TRAF	<i>Tumor Necrosis Factor Receptor-associated Factor</i> (Receptor associado ao fator de necrose tumoral)
UNICA	União da Indústria da Cana-de-açúcar

UniProt	<i>Universal Protein Resource</i>
UR	<i>Up regulated (Induzido)</i>
WTSI	<i>Wellcome Trust Sanger Institute</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>18</b>
1.1	OBJETIVOS.....	20
1.1.1	<b>Objetivo geral.....</b>	<b>20</b>
1.1.2	<b>Objetivos específicos.....</b>	<b>20</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA.....</b>	<b>21</b>
2.1	A CANA-DE-AÇÚCAR.....	21
2.1.1	<b>Importância econômica.....</b>	<b>22</b>
2.1.2	<b>Melhoramento da cana-de-açúcar.....</b>	<b>25</b>
2.2	ESTRESSE ABIÓTICO.....	26
2.2.1	<b>Déficit hídrico.....</b>	<b>27</b>
2.3	REGULAÇÃO DA EXPRESSÃO GÊNICA.....	29
2.3.1	<b>Modificações de histonas.....</b>	<b>30</b>
2.3.1.1	Histona acetiltransferases.....	31
2.3.1.2	Histona desacetilases.....	33
2.3.1.3	Histona metiltransferases.....	34
2.3.1.4	Histona demetilases.....	37
2.3.2	<b>Complexos de remodelamento da cromatina.....</b>	<b>38</b>
2.4	TRANSCRIPTÔMICA.....	40
2.4.1	<b>Ferramentas de bioinformática aplicadas ao estudo de genes estresse-responsivos em plantas.....</b>	<b>44</b>
2.4.1.1	Bancos de dados biológicos.....	45
2.4.1.2	Alinhamento de sequências e montagem de transcriptomas.....	46
2.4.1.3	Anotação funcional.....	47
2.4.1.4	Desafios da bioinformática.....	49
<b>3</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>51</b>
3.1	MATERIAL VEGETAL, ENSAIO DE DÉFICT HÍDRICO E BIBLIOTECAS HT-SuperSAGE.....	51
3.2	ANOTAÇÃO DE TRANSCRITOS CoREG, IDENTIFICAÇÃO DE DOMÍNIOS FUNCIONAIS E ANÁLISE <i>IN SILICO</i> DA EXPRESSÃO.....	52
3.3	ANÁLISE DE ENRIQUECIMENTO GO E DE FTs.....	52
3.4	MAPEAMENTO <i>IN SILICO</i> DE <i>LOCi</i> CoREG EM GENOMAS DE GRAMÍNEAS.....	53

3.5	EXTRAÇÃO DE RNAs E SÍNTESE DE cDNAs.....	53
3.6	DESENHO DE <i>PRIMERS</i> , EFICIÊNCIA DE AMPLIFICAÇÃO E ANÁLISES DE qPCR.....	54
<b>4</b>	<b>RESULTADOS.....</b>	<b>55</b>
4.1	ESTs CoREG E DISTRIBUIÇÃO EM FAMÍLIAS.....	55
4.2	CoREGs ASSOCIADOS A <i>unitags</i> HT-SuperSAGE.....	56
4.3	ENRIQUECIMENTO DE TERMOS GO.....	63
4.4	ENRIQUECIMENTO DE FTs.....	67
4.5	GENÔMICA ESTRUTURAL: PROSPECÇÃO DE <i>LOC</i> / ASSOCIADOS A CoREGs EM GENOMAS DE GRAMÍNEAS.....	69
<b>4.5.1</b>	<b>Identificação de CoREGs ortólogos em gramíneas.....</b>	<b>71</b>
4.6	VALIDAÇÃO DA EXPRESSÃO GÊNICA POR RT-qPCR.....	74
<b>5</b>	<b>DISCUSÃO.....</b>	<b>76</b>
5.1	PANORAMA DA EXPRESSÃO POR qPCR DE ALVOS CoREGs E PROVÁVEIS ATUAÇÕES EM PLANTAS SOB ESTRESSE.....	79
<b>6</b>	<b>CONCLUSÕES.....</b>	<b>85</b>
	<b>REFERÊNCIAS.....</b>	<b>86</b>
	<b>ANEXOS.....</b>	<b>99</b>
	ANEXO A - CURVAS DE <i>MELTING</i> MOSTRANDO OS PICOS DE CADA ALVO AVALIADOS EM CDNAs DE RAÍZES DE ACESSOS DE CANA-DE-AÇÚCAR SUBMETIDOS A SUPRESSÃO DE REGA POR 24 H.....	99
	ANEXO B – CURVAS PADRÃO (1:1, 1:10, 1:100, 1:1000, 1:10000) DE CADA ALVO AVALIADOS EM RAÍZES DE ACESSOS DE CANA-DE-AÇÚCAR SOB SUPRESSÃO DE REGA POR 24 H...	

## 1 INTRODUÇÃO

A cana-de-açúcar (*Saccharum* spp.) é uma gramínea com elevada capacidade de armazenar sacarose em seus colmos, tornando-a importante componente da economia de mais de 100 países. O Brasil destaca-se como o maior produtor mundial dessa cultura, bem como o principal processador e produtor de suas principais *commodities* (açúcar e etanol). No entanto, a produtividade da cultura é fortemente ameaçada pela seca, estresse abiótico que acarreta perdas consideráveis ao redor do mundo.

As plantas adquiriram mecanismos que as permitem tolerar estresses ambientais, por processos que envolvem mudanças adaptativas ou morfológicas que visam minimizar os efeitos deletérios oriundos destas condições. Dessa forma, a partir da percepção de condições adversas, complexas redes de regulação gênica (*gene regulatory networks*, GRNs), que são compostas por diferentes regulações genéticas e epigenéticas, vão acarretar mudanças na expressão de fatores de transcrição (FTs), na distribuição de nucleossomos, nas modificações de histonas, na metilação do DNA, e nos RNAs não codificadores de proteínas (*non-protein-coding* RNAs, npcRNAs), conduzindo a importantes rotas que respondem a estresses abióticos

Os FTs são os genes reguladores da expressão gênica mais conhecidos. Uma outra classe de reguladores da expressão gênica é composta de proteínas que atuam indiretamente na transcrição de seus genes-alvo (sem se ligar aos DNAs, nos elementos regulatórios dos promotores). Essas proteínas são reconhecidas como corretores (CoREGs), controlando a expressão gênica como co-ativadores ou co-repressores, por mecanismos de interação proteína-proteína em interações com FTs ou restringindo/liberando o acesso ao DNA, como modificadores de histonas ou remodeladores de cromatina. Não há relatos em relação à identificação de genes das famílias CoREGs de cana-de-açúcar, bem como, suas potenciais respostas transcricionais frente à estresses abióticos.

De modo geral, os relatos em plantas cultivadas abordam um ou poucos CoREGs. Há, porém, alguns bancos de dados que possuem genes CoREGs depositados. A base GrassCoRegDB (<http://grassius.org/grasscoregdb.html>) possui genes CoREGs de cinco espécies (*Z. mays*, *Oryza sativa* L., *Sorghum bicolor* (L.) Moench, *Brachypodium distachyon* (L.) Beauv. e *Saccharum* spp.), totalizando 579

co-reguladores em 25 famílias. Para cana-de-açúcar há, porém, representantes somente para duas famílias. Outras bases informativas, mas sem genes CoREGs de *Saccharum* spp. depositados, correspondem ao PlnTFDB 3.0 (<http://plntfdb.bio.uni-potsdam.de/v3.0/>) e iTAK (<http://bioinfo.bti.cornell.edu/cgi-bin/itak/download.cgi>). Logo, estudos de caracterizações de genes CoREGs em cana-de-açúcar e identificação daqueles com expressão diferencial, em resposta a um estresse abiótico, são importantes para os programas de melhoramento, pois podem acelerar o desenvolvimento de materiais mais adequados para cultivos em condições não satisfatórias para a cultura.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo geral

Identificar e classificar genes CoREGs de cana-de-açúcar, a partir de transcritos de bibliotecas HT-SuperSAGE, oriundos de raízes de acessos contrastantes em início de supressão de rega (24 h), ou ainda de bancos de dados públicos, bem como classificar a modulação da expressão gênica e a genômica estrutural desses transcritos visando desenvolver marcadores moleculares funcionais para uso futuro em programas de melhoramento genético da cultura.

### 1.1.2 Objetivos específicos

**3.2.1** Anotar o transcriptoma disponível, quanto à presença de transcritos expressos associados aos representantes das famílias gênicas de CoREGs.

**3.2.2** Analisar *in silico* a expressão e a modulação de transcritos CoREGs para cada conjunto de acessos (tolerante e sensível), após aplicação do estresse e em relação ao respectivo controle sem estresse, visando identificar potenciais genes estresse-responsivos.

**3.2.3** Mapear *in silico loci* associados aos CoREGs em genomas de gramíneas, a fim de analisar a sintonia e colinearidade dos genes.

**3.2.4** Validar por uma segunda técnica (RT-qPCR) a expressão observada *in silico* de transcritos CoREGs diferencialmente expressos, em resposta ao estresse imposto, a partir de oligonucleotídeos iniciadores específicos.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 A CANA-DE-AÇÚCAR

A cana-de-açúcar (*Saccharum* sp.) é uma monocotiledônea pertencente à ordem Poales, família Poaceae, subfamília Panicoideae, tribo Andropogoneae, subtribo Saccharinea e gênero *Saccharum*. É uma gramínea perene, tropical ou subtropical que evoluiu sob condições de alta temperatura, forte luz solar e grande disponibilidade de água, e possui grande habilidade de acumular altos níveis de sacarose, sendo assim, uma excelente fonte de bioenergia para os países que a cultivam (RAE *et al.*, 2005; SORENG *et al.*, 2015).

Os estágios fenológicos compreendem quatro fases de desenvolvimento. 1) Brotação e emergência: referente ao desenvolvimento dos brotos e raízes; 2) Perfilhamento: emissão dos perfilhos secundários e terciários; 3) Crescimento dos colmos ou “grande crescimento”: crescimento e desenvolvimento dos perfilhos, com ganho de altura e acumulação de açúcar em níveis basais; 4) Maturação dos colmos: rápida síntese de açúcares e concentração de fotoassimilatos (BONNETT, 2013).

No que diz respeito à sua fisiologia, a cana-de-açúcar é uma gramínea de metabolismo C4 em que a primeira molécula estável oriunda da fixação do dióxido de carbono (CO<sub>2</sub>) é o ácido oxaloacetato (OAA), este, é convertido a malato ou aspartato. Além da enzima ribulose-1,5-bisfosfato carboxilase / oxigenase (Rubisco), presente nas células de bainha do feixe, as plantas C4 possuem a fosfoenolpiruvato carboxilase (PEPc) nas células do mesófilo, a PEPc tem alta afinidade pelo CO<sub>2</sub> aumentando assim a fixação desta molécula. Devido a compartimentalização espacial destas duas enzimas, a fixação de CO<sub>2</sub> é translocada como malato e aspartato nas células da bainha do feixe, onde ocorre descarboxilação com *input* de carbono no ciclo de Calvin-Benson (TAIZ e ZEIGER, 2006).

No que se refere aos recursos genéticos da cana-de-açúcar, são conhecidos três grupos: (1) cultivares tradicionais; (2) parentes selvagens; (3) cultivares modernos. As cultivares tradicionais praticamente não são cultivadas atualmente, mas têm importância acentuada como fontes de características para novos cruzamentos e como progenitores de cultivares modernas. Os parentes selvagens incluem *S.*

*spontaneum*, *S. robustum* e outras espécies dos gêneros *Erianthus*, *Miscanthus*, *Sclerostachya* e *Narenga*. As cultivares modernas são obtidas por cruzamentos entre *S. spontaneum* e cultivares tradicionais (GRIVET *et al.*, 2006).

Sendo assim, a cana-de-açúcar moderna é um híbrido pertencente ao gênero *Saccharum*, que é composto por três espécies *S. spontaneum*, *S. robustum* e *S. officinarum* e três subespécies *S. sinense*, *S. barberi* e *S. edule* constituindo o que se denomina de ‘complexo *Saccharum*’ (MUKERJEE, 1957). O gênero tem como características o número cromossômico desbalanceado e altos níveis de poliploidia, dificultando a determinação de relações taxonômicas (AMALRAJ e BALASUNDARAM, 2006).

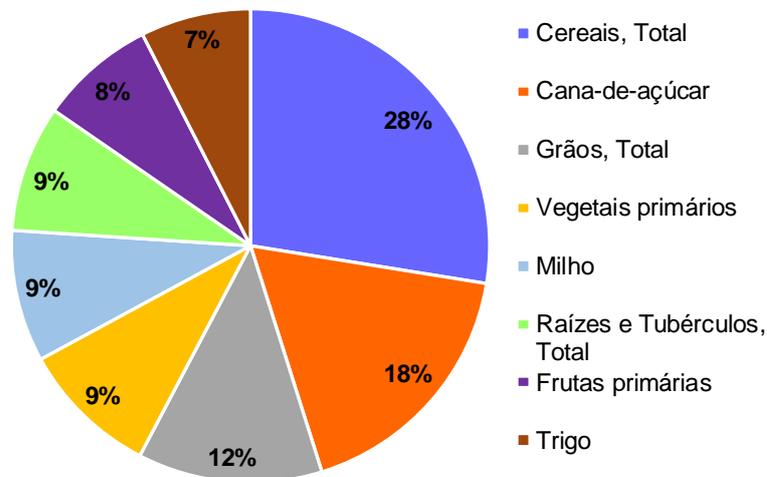
A origem da cana-de-açúcar ocorreu no Sudeste da Ásia, incluindo a Índia, Nova Guiné e algumas ilhas da Melanésia, com domesticação provavelmente ocorrendo na Nova Guiné. Hipóteses baseadas em citologia, morfologia e sobreposição de distribuição geográfica indicam que os gêneros: *Erianthus*, *Sclerostachya*, *Narenga*, e *Miscanthus* também contribuíram para a evolução da cana-de-açúcar moderna (DANIELS e ROACH, 1987).

### **2.1.1 Importância econômica**

A cana-de-açúcar é um importante componente da economia de mais de 100 países tropicais e subtropicais, conhecida por ser altamente eficiente no sentido energético (WACLAWOVSKY *et al.*, 2010). Seu valor econômico é primariamente atribuído a seus dois principais produtos – o açúcar e o biocombustível etanol. Três atributos fazem da cultura uma excelente *commoditie*: alta produtividade; uso eficiente de insumos agrícolas (água, fertilizantes etc) e a possibilidade de ser localmente processada (MOORE *et al.*, 2013).

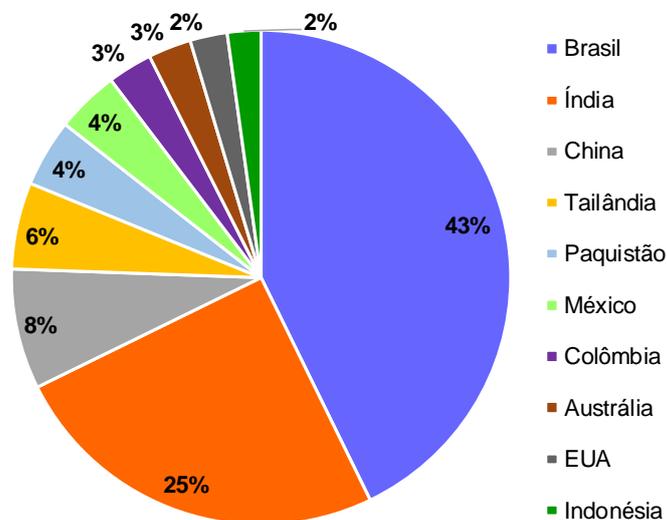
A produção mundial média da cana só perde para todo o conjunto de cereais cultivados – milho, arroz e trigo, ocupando uma área de 26,8 milhões de hectares, perfazendo cerca de 2% da área total do planeta (FAOSTAT, 2017). No que diz respeito a quantidade de calorias fornecidas aos humanos, a cana-de-açúcar ocupa o terceiro lugar propiciando 152 kcal por pessoa por dia, antecedida pelo trigo (530 kcal) e pelo arroz (533 kcal) (MOORE *et al.*, 2013).

Figura 1 - A média de produção (1994 – 2017) das culturas mais produtivas (milhões de toneladas). Fonte FAOSTAT (2017).



O Brasil desponta como o maior produtor mundial de cana-de-açúcar (Figura 2), gerando 75,2 toneladas por hectare e cobrindo uma área de 10.226.205 milhões de hectares, sendo a produção concentrada nas regiões Centro-Sul e Nordeste, assim como lidera a produção de etanol junto com os EUA, e sendo também o responsável por cerca de 20% da produção de açúcar mundial (FAOSTAT, 2017; UNICA, 2020). A alta demanda por etanol proveniente de fontes renováveis somado as condições edafoclimáticas que são propícias a cultura, permite que o Brasil ocupe lugar de destaque mundial na exportação dessa *commodity*, proporcionando cada vez mais relevância para o agronegócio brasileiro (CONAB, 2018).

Figura 2 - Média da produção (1994 – 2017) da cana-de-açúcar entre os 10 países mais produtores. Fonte FAOSTAT (2017).



A estimativa de produção para a safra 2019/20 é de 642,7 milhões de toneladas, acréscimo de 3,6% em relação à safra anterior. Na região Sudeste (maior produtora nacional) a área colhida deverá ser inferior, em razão das adversidades climáticas enfrentadas pela região que trazem estimativas de diminuição de 3,1% na sua produtividade. A região Nordeste é a segunda maior produtora e apresentou acréscimo de 12,6% em comparação com a safra passada. A região Centro-Oeste apresentou crescimento de 2,6% em relação à safra anterior, a região Sul apresentou queda de 5,7% na área colhida e a região Norte exibiu leve aumento de área colhida (CONAB, 2019).

Além da produção de etanol e açúcar, as unidades sucroalcooleiras têm objetivado elevar sua eficiência, na geração de energia elétrica, auxiliando no aumento da oferta e redução dos custos, contribuindo assim para o aumento da sustentabilidade do setor (CONAB, 2019). A bioeletricidade gerada a partir da cana-de-açúcar corresponde a quarta mais importante na matriz energética brasileira, mesmo com aproveitamento de apenas 15% do potencial com estimativa de crescimento de 50% até 2027 (UNICA, 2020).

A crescente demanda por etanol de primeira geração requer alta produção de matéria-prima, levando a preocupações sobre produção de alimento *versus* combustível e um aumento no uso da terra (SCARPARE *et al.*, 2016). Os biocombustíveis de segunda geração poderiam evitar muitas dessas preocupações, uma vez que dependem de recursos biológicos não alimentares, como os lignocelulósicos. Esses materiais lignocelulósicos são relativamente baratos e estão disponíveis em grandes quantidades, sendo o bagaço da cana (resíduo industrial da extração do caldo) o material lignocelulósico mais examinado para geração de etanol de segunda geração (BEZERRA e RAGAUSKAS, 2016; SOCCOL *et al.*, 2010).

Além do etanol de segunda geração, o bagaço da cana é utilizado para gerar outros produtos de valor agregado, como o xilitol, enzimas industriais, ácidos orgânicos, antibióticos e pigmentos. O xilitol tem elevada demanda na alimentação – como um substituto do açúcar, sendo especialmente importante para pacientes diabéticos; e na indústria farmacêutica devido à suas propriedades anti-carcinogênicas e na remineralização de dentes (PRAKASHAM *et al.*, 2009).

### 2.1.2 Melhoramento da cana-de-açúcar

Nos últimos 50 anos a cana-de-açúcar teve sua produtividade aumentada em cerca de 40%, este resultado é atribuído aos programas de melhoramento. O melhoramento clássico apresenta como desvantagem o longo período para obtenção de uma variedade mais produtiva (8-12 anos). Logo, ferramentas biotecnológicas são fundamentais para contornar os obstáculos do melhoramento clássico, sobretudo quando se trata de cana-de-açúcar, uma planta complexa devido a seus altos níveis de ploidia (aneuploidia e poliploidia) (DE MORAIS *et al.*, 2014).

Uma ampla gama de ferramentas biotecnológicas foi empregada em análises genéticas com o intuito de ampliar o conhecimento existente sobre a estrutura do genoma da cana. Como exemplo de um destes estudos, podemos citar o SUCEST (*Sugarcane EST Project* - <http://sucest-fun.org/>) que disponibilizou a maior coleção de ESTs de cana-de-açúcar do mundo, aproximadamente 238.000, oriundas de várias bibliotecas de cDNA (PINTO *et al.*, 2004). Vários estudos de expressão gênica têm sido conduzidos através de uma variedade de plataformas, com o objetivo de identificar genes específicos de um tecido (raízes, folhas, caules) associados a vários traços de interesse como: conteúdo de sacarose, síntese da parede celular e resposta a seca (RODRIGUES *et al.*, 2009; MANNERS e CASU, 2011).

No Brasil existem três programas voltados ao melhoramento da cultura: (1) Rede Interuniversitária para o Desenvolvimento do Setor Sucroalcooleiro (RIDESA), o maior programa de melhoramento de cana-de-açúcar responsável por mais de 65% das variedades cultivadas no país (RIDESA, 2020). É formado por um consórcio entre universidades federais (UFPR, UFRRJ, UFSCar, UFS, UFV, UFAL, UFG, UFMT, UFRPE e UFPI). (2) Centro de Tecnologia Canavieira (CTC), surgiu no fim da década de 1960, a partir de uma iniciativa entre usinas de Piracicaba (SP), com o intuito de investir do desenvolvimento de variedades mais produtivas. O programa detém mais de quatro mil variedades, sendo o maior banco de germoplasma de cana-de-açúcar do mundo localizado em Camamu (BA) (CTC, 2019). (3) Instituto Agrônomo de Campinas (IAC), pioneiro da pesquisa agrícola no país, foi fundado em 1892 pelo cientista Franz Dafert que avaliou 42 variedades de cana (*Saccharum officinarum*) sob

duas condições de cultivo. O programa é o responsável por 39 variedades de cana-de-açúcar, principalmente associadas ao alto teor de sacarose (IAC, 2019)

Os objetivos futuros para o melhoramento da cultura incluem o desenvolvimento de variedades altamente produtivas (tolerantes a seca, com elevado conteúdo de sacarose e produções superiores de etanol e biomassa). Da mesma forma, os melhoristas também objetivam uma planta com alto conteúdo de fibras e parede celular fácil de ser quebrada para obtenção de etanol, e uma planta cada vez mais eficiente que requeira cada vez menos insumos e tenha baixo impacto ambiental (DE MORAIS *et al.*, 2014).

## 2.2 ESTRESSE ABIÓTICO

As plantas necessitam de energia (luz), água, carbono e nutrientes minerais para crescer. Um estresse é definido como qualquer fator que restringe o desenvolvimento normal das plantas que, ao longo da evolução, desenvolveram mecanismos que as permitem sobreviver e crescer sob condições adversas. As respostas das plantas a estresses abióticos são complexas e dinâmicas, podendo ser reversíveis – mudanças morfológicas ou fisiológicas transientes, como por exemplo a diminuição da fotossíntese nas horas mais quentes do dia para conservar água durante elevada evapotranspiração ou irreversíveis – traços adaptativos, como por exemplo a presença de glândulas de sal em halófitas (MOORE *et al.*, 2013).

Os estresses abióticos que estão presentes no ambiente natural das plantas incluem a seca, salinidade do solo, calor, frio, deficiência de nutrientes e metais pesados presentes no solo. A resposta molecular das plantas a estresses abióticos envolve interações com uma série de vias, espécies reativas de oxigênio (ROS) e nitrogênio (RNS) são os sinais iniciais em muitos destes estresses, os quais, são responsáveis por regulação de genes e modificação da atividade de enzimas (MOLASSIOTIS e FOTOPOULOS, 2011).

As espécies reativas de oxigênio incluem o radical superóxido ( $O_2^{\cdot -}$ ), peróxido de hidrogênio ( $H_2O_2$ ), radical hidróxido ( $OH\cdot$ ) e oxigênio singlete ( $^1O_2$ ). Sob estresse abiótico, ocorre o que se chama de estresse oxidativo pela acumulação de ROS, provocando danos oxidativos e até morte celular (Miller *et al.*, 2008). Os radicais livres

de óxido nítrico (NO) incluem trióxido de dinitrogênio ( $N_2O_3$ ), dióxido de nitrogênio ( $NO_2$ ), peroxinitrito ( $ONOO\cdot$ ), S-nitrosotióis (RSNOs), S-nitrosoglutationa (GSNO) etc. Há fortes indícios de que as plantas utilizam ambas moléculas ROS e RNS como sinais de transdução durante processos celulares básicos (MOLASSIOTIS *et al.*, 2011).

Os fitorreguladores são igualmente importantes na resposta a estresses abióticos, sendo os dois mais consideráveis o ácido abscísico (ABA) e o etileno (GODA *et al.*, 2008). O ABA é considerado o componente central na resposta a fatores ambientais, principalmente quando se trata de estresse osmótico. A sinalização de ABA pode ocorrer de forma rápida, não envolvendo atividade transcricional (ex.: controle da abertura estomatal) ou respostas mais lentas que regulam o crescimento e germinação (CRAMER, 2010; KIM *et al.*, 2010).

As respostas das plantas a estresses abióticos envolvendo ABA podem ocorrer por vias ABA-dependente e ABA-independente. Na maneira ABA-dependente, os promotores dos genes que respondem ao ABA contém o elemento ABA responsável (ABRE) e os fatores de transcrição que se ligam a esta região (AREB/ABF) regulam a resposta ao estresse. Na via ABA-independente, as proteínas de ligação ao elemento responsivo de desidratação (DREB1) e DREB2 se ligam ao elemento DRE/CRT encontrado nos promotores, e regulam assim a resposta ao estresse abiótico (NAKASHIMA *et al.*, 2014).

### **2.2.1 Déficit hídrico**

A UNESCO considera a escassez de água como ameaça chave do século XXI. As mudanças climáticas provocadas por atividades humanas - pela mudança no regime de chuvas, inundações, seca e redistribuição geográfica de doenças - somado ao crescimento da demanda mundial por alimento, constituem graves ameaças a segurança alimentar global (RYAN, 2011). As plantas são constituídas de até 95% de água e aproximadamente 97% da água absorvida pelas é perdida para a atmosfera pelo processo de transpiração e apenas 2% é utilizado para a manutenção em processos metabólicos, como a fotossíntese. Dessa forma, não é espantoso que a limitação de água tenha sérias consequências para os vegetais (ZINGARETTI *et al.*, 2012).

A seca é o principal estresse abiótico responsável pela perda de produtividade em todo mundo, tornando urgente a necessidade de desenvolvimento de culturas cada vez mais eficientes no uso de água e pode ser considerado como um desvio das condições ótimas de vida (BODNER *et al.*, 2015; ZINGARETTI *et al.*, 2012). Em vegetações naturais, a resistência a seca é um processo evolutivo que caminha em direção ao sucesso reprodutivo, no que diz respeito a plantas cultivadas isto está diretamente relacionado à produtividade e mecanismos associados à essa resistência em populações naturais (ex.: diminuição da área foliar) são altamente eficientes, porém, incompatíveis com altas produtividades (BODNER *et al.*, 2015).

Os mecanismos de tolerância ao déficit hídrico abrangem resumidamente os conceitos de escape, evitar a desidratação e tolerância. O escape se refere ao fechamento do ciclo de vida da planta, antes da chegada do estresse (KOOYERS, 2015). Este conceito não se aplica a cana-de-açúcar pelo fato de ser uma gramínea perene. Evitar a desidratação engloba mecanismos que asseguram a hidratação celular, através da baixa condutância estomatal durante o estresse. A tolerância especifica mecanismos que asseguram que a planta tolere e mantenha suas funções sob déficit hídrico (BLUM, 2005).

As fases do desenvolvimento da cana-de-açúcar mais susceptíveis a este estresse abiótico são as de perfilhamento e a de crescimento dos colmos, sendo os órgãos mais afetados as folhas e as hastes. As respostas fisiológicas e morfológicas da cana dependem da duração e intensidade do estresse, do tecido afetado e do genótipo (ZINGARETTI, *et al.*, 2012). As mais frequentes respostas da cana ao déficit hídrico são: fechamento estomatal, rolamento das folhas e inibição do crescimento da haste e das folhas, as respostas fisiológicas incluem a diminuição da condutância estomatal ( $g_s$ ), da concentração interna de  $CO_2$  ( $C_i$ ) e da taxa de transpiração ( $E$ ) (INMAN-BAMBER e SMITH, 2005).

A cana-de-açúcar está entre as culturas que produz maior quantidade de biomassa por unidade de área e demanda elevados níveis de água durante seu crescimento, sendo considerada como altamente sensível ao déficit hídrico, podendo haver perdas de até 60% da sua produtividade (GENTILE *et al.*, 2015). Devido à crescente demanda por açúcar e etanol, as áreas de cultivo da cana são expandidas para regiões de baixa precipitação, havendo a necessidade de irrigação parcial ou

total dessas áreas, processo que já ocorre em países como a África do Sul e Austrália em que aproximadamente 60% da cana é crescida em áreas irrigadas, elevando consideravelmente o custo da produção (ZINGARETTI *et al.*, 2012).

No Brasil, a cana é produzida sob regime sequeiro, ou seja, as plantações ocupam áreas consideradas como terras áridas com limitada disponibilidade de água, mas a irrigação é uma alternativa para evitar perdas na produtividade nessas regiões. Contudo, para atender a demanda por açúcar e etanol a necessidade de irrigação ameaça a integridade dos recursos hídricos (SCARPARE *et al.*, 2016).

Os estudos acerca da biologia do estresse proveniente de estudos genéticos, agrônômicos e de biologia molecular em várias culturas, incluindo a cana-de-açúcar, melhoram o nosso entendimento do crescimento e desenvolvimento da planta sob estresse e está proporcionando um grande impulso ao desenvolvimento de estratégias biotecnológicas para produzir variedades de cana tolerantes ao estresse hídrico e comercialmente úteis (ZINGARETTI *et al.*, 2012; RICHARDS, 2006).

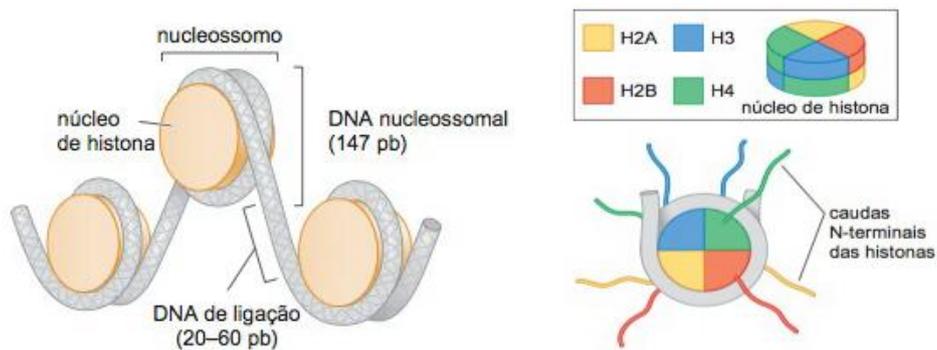
### 2.3 REGULAÇÃO DA EXPRESSÃO GÊNICA

Os genes estresse-responsivos a seca são classificados em duas categorias: genes que codificam proteínas funcionais e genes que codificam proteínas regulatórias. O grupo de genes funcionais abrange proteínas tais como as chaperonas, aquaporinas, transportadores de membrana, enzimas dentre várias outras. O grupo de genes regulatórios engloba proteínas responsáveis pela sinalização e fatores de transcrição (SHINOZAKI e YAMAGUCHI-SHINOZAKI, 2007).

Em organismos eucariotos, o DNA (ácido desoxirribonucleico) genômico é altamente compactado em uma estrutura denominada cromatina. Esta compactação permite que a molécula seja comprimida no núcleo da célula, porém, impossibilita qualquer processo biológico que necessita de acesso ao DNA tais como: replicação, transcrição, reparo do DNA, recombinação. Sendo assim, a regulação da transcrição não é determinada apenas pela informação presente nas sequências de DNA, mas é também mediada por modificações dinâmicas na estrutura da cromatina (CHEN e TIAN, 2007).

A cromatina é compactada por nucleossomos, cada um consistindo de aproximadamente 146 pb de DNA em torno de um octâmero de histonas (duas moléculas de cada um dos núcleos: H2A, H2B, H3 e H4, sendo os domínios globulares destas histonas essenciais para a manutenção da integridade estrutural do nucleossomos (LUGER *et al.*, 1998). H2A, H2B, H3 e H4 são histonas do núcleo e a histona H1 liga-se ao DNA de ligação. As histonas H3 e H4 formam um único tetrâmero e H2A e H2B formam dois dímeros e associação desses blocos com o DNA ocorre de maneira ordenada, primeiro o tetrâmero H3/H4 e depois os dímeros H2A/H2B (Figura 3).

Figura 3 - DNA compactado em nucleossomos. (a) Esquema de empacotamento e organização dos nucleossomos, evidenciando o DNA enrolado ao redor de um núcleo de histonas. (b) A montagem do nucleossomo, evidenciando as diferentes histonas que compõem o núcleo.



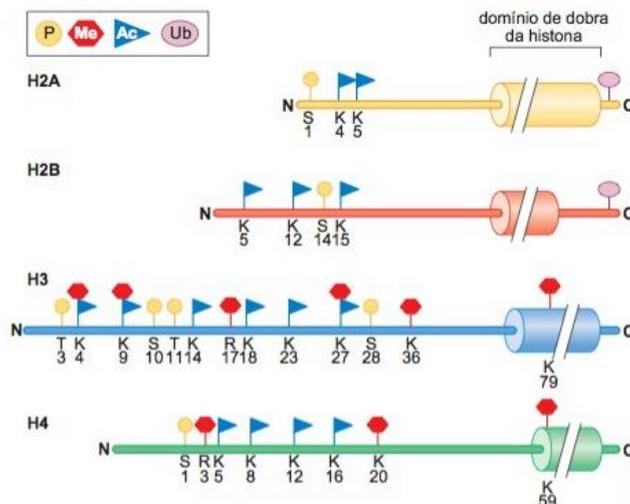
A estrutura e função da cromatina são regulados por uma variedade de mecanismos epigenéticos (mudanças que alteram a atividade do DNA sem que haja modificação da sequência nucleotídica) que incluem as modificações de histonas, a metilação do DNA, complexo de remodelamento da cromatina ATP-dependente, substituição de variantes de histonas e regulação por RNAs não codificantes. Juntos, estes processos são implicados por modular a estrutura da cromatina e, conseqüentemente, a atividade gênica (LIU *et al.*, 2010).

### 2.3.1 Modificações de histonas

As modificações de histonas têm papel fundamental na regulação da expressão gênica. As modificações pós-traducionais tais como acetilação, metilação, fosforilação, que ocorrem nas caudas N-terminais das histonas, são responsáveis pela

interação com DNA e assim facilitando a montagem da cromatina (TURNER, 2002) (Figura 4). A combinação de variantes de histonas e da variedade de modificações pós-traducionais que as mesmas experimentam, é considerada um “código de histonas” que provê sinais para recrutamento de proteínas específicas, as quais, afetam a regulação transcricional (TURNER, 2002).

Figura 4 - Esquema das modificações pós-traducionais nas caudas de histonas. Para fins de simplificação estão ilustrados apenas sítios de acetilação, metilação, fosforilação e ubiquitinação.



A acetilação de histonas ocorre nos resíduos de lisina (K) e arginina (R) numa reação catalisada por acetiltransferases (HATs), que é comumente relacionada a ativação gênica, enquanto que a desacetilação catalisada por desacetilases (HDACs) é frequentemente associada ao decréscimo da atividade gênica (CHEN e TIAN, 2007; FU *et al.*, 2007). A metilação das caudas também tem papel fundamental na regulação de vários processos ao longo do desenvolvimento, estando envolvido também no silenciamento de sequências repetitivas visando manter a estabilidade do genoma. As marcas de metilação são adicionadas aos resíduos de lisina e arginina às caudas de histonas através de enzimas com atividade de histonas metiltransferases (HMTs) e removidas pelas histonas demetilases (HMDs) (LIU *et al.*, 2010).

### 2.3.1.1 Histona acetiltransferases

As histonas acetiltransferases catalisam a transferência da porção acetil do grupo Acetil-coA para o grupo  $\epsilon$ -amino dos aminoácidos presentes nas caudas das histonas. As mudanças rápidas e reversíveis na acetilação de histonas iniciam papel chave na modificação da cromatina, conduzindo alterações na expressão de genes e afetando assim processos biológicos em resposta à sinais externos e internos, tais como crescimento, desenvolvimento, diferenciação celular e estresses bióticos e abióticos (LUSSER *et al.*, 2001).

As caudas amino-terminais das histonas possuem alto conteúdo de lisina e arginina, sendo assim de natureza altamente básica. Inicialmente foi proposto que os resíduos de lisina quando acetilados, neutralizam a carga positiva das caudas das histonas diminuindo sua afinidade ao DNA, que possui carga negativa, permitindo que reguladores transcricionais tenham acesso ao DNA (KUO *et al.*, 1998; LUGER e RICHMOND, 1998). Porém, evidências mais recentes sugerem que a acetilação também diminui a interação entre nucleossomos vizinhos, prevendo a compactação dos nucleossomos (SHAHBAZIAN e GRUNSTEIN, 2007).

Em organismos eucariotos, a acetilação de histonas é catalisada por cinco famílias distintas, compreendendo a família p300/CREB (proteína de ligação ao elemento c-AMP responsivo), a família CBP, família (TAF)<sub>II</sub>250 (fator de associação à proteína de ligação à TATA), família GNAT (controle geral não repressível-5 relacionado à acetiltransferase N-terminal), a família MYST (MOZ, Ybf2/Sas3, Sas2, e Tip60) e a família de HATs relacionada à hormônios nucleares (PANDEY *et al.*, 2002).

O envolvimento de HATs na resposta a estresses abióticos tem sido bem relatado. ADA2b é provavelmente envolvido na resposta a estresse abiótico, visto que, os mutantes *ada2b* exibiram hipersensibilidade ao estresse salino e ao ABA (HARK *et al.*, 2009). A HAT GCN5 se mostrou crucial na regulação da termotolerância em *Arabidopsis*, onde sua perda de função no mutante *gcn5* submetido a tratamentos de 38 e 45 °C, resultou na repressão dos fatores de transcrição HSFA3 e HSFA3, acarretando em sérios defeitos em termos de tolerância a elevadas temperaturas (HU *et al.*, 2015).

Em *Beta vulgaris* e *B. maritima*, um ensaio de imunoprecipitação da cromatina (ChIP) em plantas controle e sob estresse salino (250 e 500 mM de NaCl), identificou que um gene que codifica uma peroxidase (POX) estava transcricionalmente ativo, isto se correlacionou com elevados níveis de acetilação em H3K4 e H3K27 (YOLCU *et al.*, 2016). Em plantas de arroz submetidas ao déficit hídrico, foi observado um aumento da expressão de quatro HATs (*OsHAC703*, *OsHAG703*, *OsHAF701* e *OsHAM701*), que estava associado com o acréscimo da acetilação em H3 (lisinas 9, 18 e 27) e H4 (lisina 5) (FANG *et al.*, 2014).

### 2.3.1.2 Histona desacetilases

A desacetilação envolve a remoção de grupamentos acetil de resíduos específicos de lisina do core de histonas, reação que é catalisada por histonas desacetilases com o intuito de restaurar as cargas positivas destes resíduos. A desacetilação está associada a uma cromatina mais compacta, bloqueando a acessibilidade de fatores de transcrição e conduzindo, portanto, ao silenciamento de genes (HOLLENDER e LIU, 2008). As histonas desacetilases são atualmente agrupadas em três famílias, as RPD3/HDA1 (proteína-3 de dependência reduzida de potássio/histona desacetilase-1), SIR2 (proteína reguladora do silenciamento de informação-2) e H2D (histona desacetilase-2), que são desacetilases específicas de plantas (LUSSER *et al.*, 2001).

O papel das desacetilases em uma variedade de processos biológicos tem sido evidenciado em um elevado número de estudos. Plantas de *Arabidopsis* super-expressando HDA19, exibiram acréscimo de resistência ao fungo *Alternaria brassicicola*, a expressão de HDA19 e HDA6 se mostrou induzida pelo ataque do patógeno ou por ferimento, sendo proposto que HDA19 pode interagir com fatores de transcrição WRKY para regular as respostas de defesa basal em plantas (KIM *et al.*, 2008). Há evidências que indicam que HDA19 e HDA6 possuem papel na modulação da germinação e resposta ao estresse salino, bem como expressão gênica induzida por ABA e estresse salino em *Arabidopsis*.

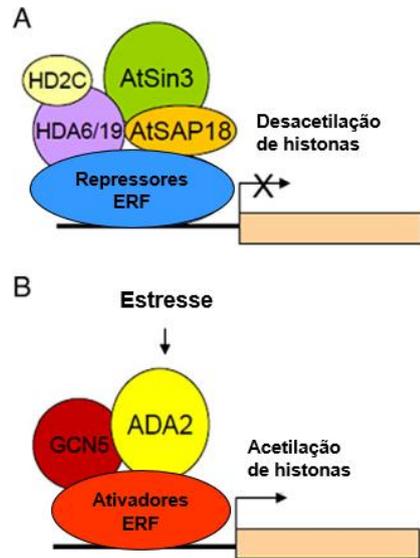
As plantas HDA6 mutantes, *axe1-5*, exibem fenótipo mais sensível ao ABA, comparado com plantas selvagens, a expressão de genes responsivos a estresses abióticos e ao ABA foi menor em *axe1-5* e os mutantes *hda19-1* se mostraram

hipersensíveis ao ABA e sal, com expressão dos genes ABA responsivos reduzida (CHEN *et al.*, 2010a; CHEN *et al.*, 2010b). A HDAC da família HD2, HD2C, também se mostrou envolvida na resposta ao ABA e ao estresse salino em plantas de *Arabidopsis*, pois sua super-expressão conferiu um fenótipo ABA-insensível e aumento da tolerância a seca e salinidade (SRIDHA e WU, 2006).

Em *Arabidopsis*, HDA19 tem sido associada a repressores transcricionais ERF, AtERF3, AtERF4 e AtERF7, para regular a expressão gênica em resposta a estresses abióticos. Foi observado que AtERF7 interage com AtSin3 e HDA19, formando um complexo repressor. AtERF3 e AtERF4 podem interagir com AtSAP18 (ortólogo de SAP18 em humanos, uma subunidade do complexo HDAC). Juntos AtERF3, AtERF4, AtERF7, AtSin3 e AtSAP18 são propostos para formar um complexo que regula a expressão gênica em resposta ao ABA e estresses abióticos (SONG e GALBRAITH, 2006).

Os repressores ERF podem reprimir a transcrição através de modificações na cromatina via atividade de HDACs. Enquanto que os ativadores ERF, tais como CBF1, atuam estimulando a transcrição por meio do recrutamento de complexos contendo GCN5-HAT. O balanço entre ativadores e repressores ERF nos promotores de genes alvo em comum, fornece mecanismos consideráveis a respeito do controle da transcrição durante a exposição de plantas a estresses abióticos (LUO *et al.*, 2012) (Figura 5).

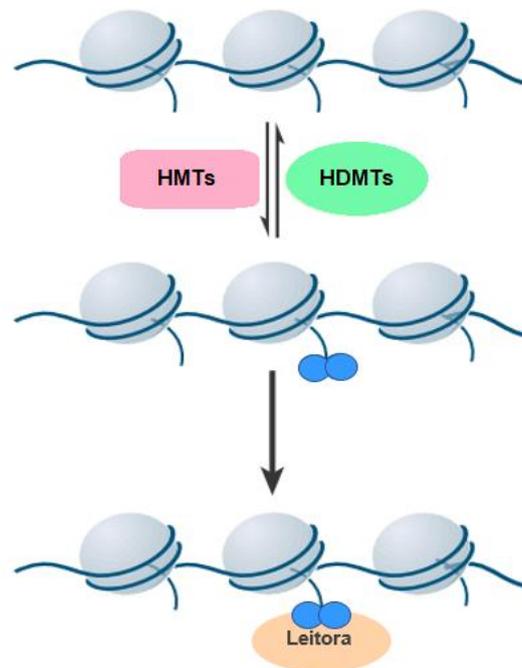
Figura 5 - Representação esquemática da ação de HATs e HDACs na regulação de respostas a estresses abióticos. Traduzido de Luo *et al.*, (2012).



### 2.3.1.3 Histona metiltransferases

A metilação de histonas, cumpre papel essencial em uma ampla gama de processos biológicos que envolvem desde a regulação transcricional até a formação da heterocromatina. Ambos aminoácidos lisina e arginina podem ser metilados, a reação é catalisada por duas enzimas distintas: histona lisina metiltransferases (HKMTs) e as proteínas arginina metiltransferases (PRMTs) (LIU *et al.*, 2010). O *status* desta metilação é interpretado por outro grupo de enzimas denominadas de 'leitoras', as quais, ligam-se especificamente a aminoácidos metilados na cauda de H3, os domínios responsáveis por esta função são PHD, chromo, WD40, Tudor, MBT e PWWP (TAVERNA *et al.*, 2007) (Figura 6).

Figura 6 - Representação esquemática do processo adição (HMTs), remoção (HDMTs) e interpretação de grupos metil adicionados a caudas de histonas. (LIU *et al.*, 2010).



A metilação das lisinas nas caudas de histonas é catalisada por um grupo de proteínas que contém os domínios SET e DOT1, podendo ocorrer nos contextos mono, di e tri-metilados e contribuem para ambas funções de ativação e repressão da cromatina. As marcas de metilação geralmente associadas com silenciamento são H3K9 e H3K27 e as de ativação H3K4 e H3K36 (BERGER, 2007).

A reação de metilação das lisinas não afeta a carga líquida dos resíduos modificados, porém, eleva a hidrofobicidade e pode alterar as interações intra ou intermoleculares ou criar novas superfícies de ligação para as proteínas do leitores que se ligam preferencialmente a domínios metilados. É considerada uma marca epigenética significativa e complexa, estando presente em regiões ativadas e inativadas da cromatina dependendo de qual resíduo de lisina está metilado e do grau de metilação (LIU *et al.*, 2010).

A metilação das argininas é catalisada pelas PRMTs, podendo ocorrer apenas nos contextos mono e di-metil (simétrica ou assimétrica) e tem sido relacionada a vários processos biológicos como a regulação da transcrição, processamento do RNA,

transporte nuclear, transdução de sinal e reparo do DNA. As argininas mais frequentemente metiladas são: Arg2, Arg8, Arg17 e Arg26 da histona H3 e Arg3 da histona H4 (KLOSE e ZHANG, 2007).

Uma análise de marcas de metilação (mono, di e tri) de H3K4 em todo o genoma de plantas de *Arabidopsis* expostas ao déficit hídrico, encontrou que H3K4me3 apresentou as mudanças mais significativas, e que as diferenças se correlacionaram positivamente com diferenças nos níveis de transcrição quando os genes foram agrupados de acordo com seus níveis de expressão, sendo que as alterações em H3K4me3 foram mais acentuadas nos genes mais diferencialmente expressos (VAN DIJK *et al.*, 2010).

Também sob condições de déficit hídrico, foi encontrado um aumento nos níveis de transcrição de *RD29A/B* bem como *NCED3* (que codifica uma enzima que atua na biossíntese do ABA), que foi acompanhado do aumento de H3K4me3. No mutante *atx1*, as mudanças transcricionais e as marcas H3K4me3 foram severamente diminuídas bem como, apresentando sintomas de perda de água mais pronunciados em comparação com as plantas selvagens (DING *et al.*, 2011).

ZONG *et al.*, 2013 investigaram o padrão de distribuição da marca H3K4me3 e sua associação com perfis de expressão no genoma do arroz (*Oryza sativa*) sob condições de déficit hídrico. Os autores associaram H3K4me3 a mudanças de expressão de alguns genes que atuam na tolerância à seca. Uma região específica do cromossomo 11, com mudanças na marca H3K4me3 após o tratamento de seca contém quatro clusters de genes desidrata, os quais, exibiram altos níveis da marca e expressão fortemente induzida pelo déficit hídrico.

Em *Arabidopsis*, a metiltransferase de arginina SKB1 também denominada de PRMT5 confere tolerância ao estresse salino, por regular a transcrição e o *splicing* pela alteração dos níveis de H4R3me2 e metilação das pequenas ribonucleoproteínas (LSM4) (ZHANG *et al.*, 2011). A metiltransferase ATX1 é responsável pela repressão do gene *WRKY70* durante a seca em plantas de *Arabidopsis*, e que os baixos níveis dos transcritos estão associados a drástica redução de H3K4me3 e aos elevados níveis de fosfatidilinositol 5-fosfato, que atuam mantendo a metiltransferase no citoplasma, suprimindo sua função no núcleo (NDAMUKONG *et al.*, 2010).

#### 2.3.1.4 Histona demetilases

As marcas de metilação em histonas são removidas pelas histonas demetilases (HDMs), contribuindo para a homeostase do processo. Há dois tipos de demetilases de histonas com mecanismos de ação distintos, (1) oxidação da amina pela demetilase específica de lisina 1 (LSD1/KDM1) e (2) hidroxilação por proteínas contendo o domínio Jumonji (JmjC) (KLOSE e ZHANG, 2007).

A reação de demetilação por LSD1/KDM1 requer um segundo fator denominado Co-REST (repressor transcricional associado a cromatina), também foi observado que a atividade deste complexo perde a eficiência em nucleossomos hiperacetilados e é estimulada pela atividade da deacetilase de histona (HDAC1), revelando assim, a relação entre as atividades demetilase e deacetilase (SHI *et al.*, 2005).

Devido a prevalência de histonas tri-metiladas, e ao fato de que as LSD1/KDM1 serem limitadas aos substratos mono e di-metilados, os pesquisadores postularam que provavelmente deveria existir uma classe adicional de histonas demetilases. As similaridades entre os domínios catalíticos da demetilase bacteriana AlkB e JmjC, demonstrou que a histona demetilase 1A contendo o domínio JmjC (JHDM1A, também conhecida como Fbx11) constitui demetilases de histona, na qual, o domínio JmjC foi considerado o *core* catalítico da enzima, porém, JmjC, não é capaz de remover grupos metil no estado tri-metil (TREWICK *et al.*, 2005).

O papel das histonas demetilases em resposta a estresses abióticos não tem sido muito estudado, mas estudos mais recentes sugerem seu envolvimento durante condições estressantes. SONG *et al.*, 2018 analisaram o crescimento e desenvolvimento de plantas de arroz que superexpressavam a demetilase OsJMJ703 e plantas nocauteadas para OsJMJ703, bem como os padrões de expressão deste gene em plantas submetidas a déficit hídrico. Os autores observaram que as plantas que não expressavam o gene da demetilase, exibiram maior tolerância ao estresse e que as plantas que superexpressavam se mostraram mais sensíveis, quando comparado com plantas selvagens.

A demetilase JMJ15 é envolvida na tolerância ao estresse salino em *Arabidopsis*. Em mutantes com ganho de função (*jmj15-1*, *jmj15-2*), foi observada uma

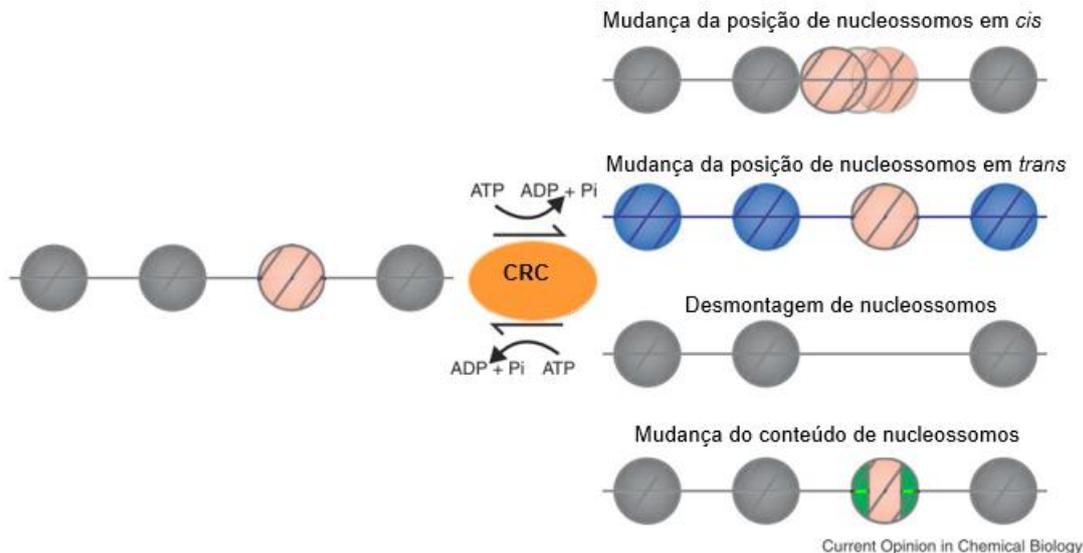
maior tolerância ao referido estresse, enquanto que o mutante com perda de função (*jmj15-3*) exibiu fenótipo sensível, quando comparado com as plantas selvagens. Nos mutantes de ganho de função, os níveis de transcritos do gene JMJ15 foi elevado e isto foi associado a repressão de genes, visto que o número de genes reprimidos se sobressaiu ao de genes induzidos, 23 e 164 respectivamente (SHEN *et al.*, 2014).

### **2.3.2 Complexos de remodelamento da cromatina**

Os complexos de remodelamento da cromatina (CRCs) utilizam a energia proveniente da hidrólise do ATP para alterar a estrutura da cromatina, e consequentemente atuando na regulação da expressão gênica eucariótica. Embora existam complexos adicionais, estes são classificados em quatro classes principais: (1) as ATPases SWI/SNF; (2) as ATPases *switch* (ISWI); (3) ATPases contendo cromodomínio e domínios similares a helicases (CHD); (4) INO80/SWR. (SUDARSANEN e WISNTON, 2000; HAN *et al.*, 2015).

A ação de vários complexos de remodelamento gera diferentes resultados, tais como a mudança na posição de nucleossomos ao longo do DNA em *cis* e em *trans*, a montagem/desmontagem de nucleossomos e a mudança dos conteúdos de histonas dentro do nucleossomo, como por exemplo a incorporação de variantes específicas de histonas (CLAPIER e CLAINS 2009) (Figura 7).

Figura 7 - Os diferentes resultados da ação de complexos de remodelamento da cromatina (CRCs).



Cada um destes complexos desempenha atividades bioquímicas únicas. Por exemplo, as classes CHD e ISWI têm funções únicas no espaçamento dos nucleossomos e na montagem da cromatina após a replicação (CORONA e TANKUN, 2004). Os complexos INO80 e SWR1 têm funções opostas na substituição de variantes de histonas, enquanto INO80 substitui H2A.Z por H2A, SWR1 catalisa a substituição de H2A por H2A.Z, sendo que a presença de H2A ou H2A.Z afeta a estabilidade do nucleossomos (KUMAR e WIDGE, 2010).

Estes complexos foram inicialmente associados a ativação transcricional, pela habilidade de desmontar ou deslizar os nucleossomos e cooperando com ativadores como a histona acetiltransferase GCN5 (COSMA *et al.*, 1999; CLAPIER e CLAINS, 2009). Recentes estudos têm evidenciado a atividade repressora deste complexo, os mecanismos que constituem o papel repressivo incluem: o favorecimento ao recrutamento de proteínas que se ligam a sequências específicas ou reguladores da cromatina com atividades repressivas (WU *et al.*, 2014); interações de longo alcance entre *enhancers* e promotores (KIM *et al.*, 2009); estabilização de nucleossomos no sítio de início da transcrição (ZHU *et al.*, 2013).

A literatura relata o envolvimento de grupos de complexos remodeladores da cromatina com respostas à estresses abióticos. BUSZEWICS *et al.*, 2016

evidenciaram a interação entre uma histona desacetilase (HD2C) e um destes complexos (BRM) em plantas de *Arabidopsis* expostas a elevada temperatura. Os autores observaram que havia um subconjunto de genes comumente regulados nos mutantes *hd2c* e *brm*, pois todos pertenciam a mesma categoria GO, sendo uma parte destes genes induzidos incluindo genes de resposta já conhecida a este estresse, tal como HSP101. Os dados apoiam a classificação de HD2C e BRM como repressores de genes responsivos ao calor e sugerem que HD2C e BRM atuam juntos no mesmo complexo proteico.

O homólogo de *Arabidopsis* SWI3D em milho (ZmCHB101) é apontado como tendo papel essencial na resposta ao déficit hídrico e estresses osmóticos. Plantas transgênicas ZmCHB101-RNAi exibiram fenótipos sensíveis a ambas condições. A análise RNA-seq identificou que ZmCHB101 tem impacto na expressão de genes responsivos a estresse osmótico, ativando especialmente genes codificantes de TFs, incluindo as famílias AP2/EREBP, WRKY e FTs *heat shock*, os quais, já possuem atuação conhecida em processos de tolerância a estresses (YU *et al.*, 2018).

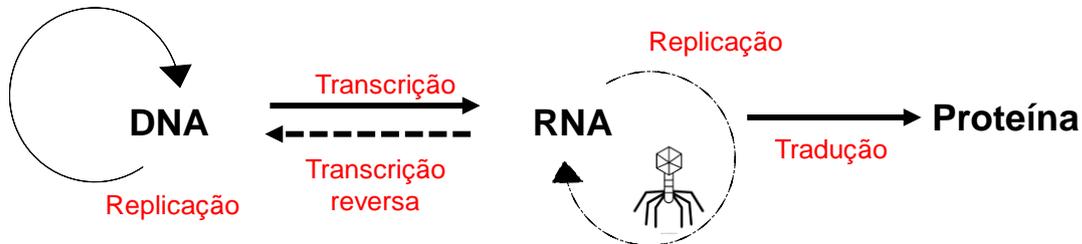
O subgrupo de SWI/SNF BRAHMA (BRM) modula a resposta ao ABA impedindo a ativação prematura de vias de resposta ao estresse durante a germinação em *Arabidopsis*, através da interação dos componentes da via de sinalização do ABA com BRM por fosforilação/defosforilação. O mutante fosfomimético BRM<sup>S1760D S1762D</sup> exibiu hipersensibilidade ao tratamento com ABA, indicando que o controle de BRM por rápida defosforilação pode ser uma característica que aumenta a tolerância a seca nestas plantas (PEIRATS-LLOBET *et al.*, 2016).

## 2.4 TRANSCRIPTÔMICA

A partir da publicação do ‘Dogma Central da Biologia Molecular’ em 1970, proposto por Francis Crick, foram indicados três princípios que regem o fluxo da informação genética: (1) a informação genética está presente no DNA; (2) a informação presente na molécula de DNA é capaz de se autorreplicar (replicação), esta mesma informação é transferida do DNA para o RNA através da transcrição e a informação contida no RNA é capaz de determinar a composição de um polipeptídeo,

através da tradução; (3) a informação nunca é transmitida da proteína para o DNA (Figura 8).

Figura 8 - Representação esquemática sintetizada do Dogma Central da Biologia Molecular.



O desenvolvimento das tecnologias de sequenciamento em larga escala de ácidos nucleicos (DNA e RNA), permitiu uma ampla visão dos elementos que atuam no fluxo da informação genética. A partir da vasta quantidade de conhecimento gerado por estes métodos, nascia a transcriptômica, que é definida como o estudo do conjunto de RNAs de um tipo celular específico ou de um tecido em determinado estágio de desenvolvimento ou condição fisiológica, que inclui os RNAs mensageiros, RNAs transportadores, RNAs ribossomais e outros RNAs não codificantes (DONG e CHEN, 2013).

Quatro razões fazem da transcriptômica um campo de pesquisa inspirador na era pós-genômica. (1) As análises em transcriptômica refletem a dinâmica da expressão de um genoma; (2) A transcriptômica apoia os estudos no campo da proteômica; (3) Estudos funcionais e estruturais dos RNAs não codificantes de proteínas (ncRNA) ampliam o propósito da transcriptômica; (4) As inovações metodológicas, especialmente das tecnologias de sequenciamento de nova geração (NGS), permitiram um maior nível de rendimento e resolução de estudos de transcriptomas e geraram mais dados com significados biológicos (LOCKHART e WINZELER, 2000).

O transcriptoma é altamente dinâmico e responde rápida e dramaticamente a desordens ou mesmo a eventos celulares comuns, como a divisão celular por exemplo. O entendimento de quando, onde e até que ponto um gene é expresso é crucial para entender a atividade e as funções biológicas exercidas a partir disto. Adicionalmente, as mudanças nos padrões de expressão podem fornecer indícios de

mecanismos regulatórios e vias bioquímicas mais amplas (LOCKHART e WINZELER, 2000).

SHIMKETS (2004) classifica as técnicas de análise de transcriptomas em três categorias principais: (i) Metodologias baseadas na visualização do transcrito – a geração de transcritos marcados pode ser quantificada e comparada entre as diferentes amostras; (ii) Metodologias baseadas em hibridização – explora a complementariedade das bases nitrogenadas e podem ser subdivididas em mais duas categorias: hibridização em superfície sólida e hibridização em solução; (iii) Metodologias baseadas no sequenciamento e contagem de transcritos.

Dentre as tecnologias mais utilizadas para análises de expressão gênica, destacam-se as técnicas: *complementary DNA (cDNA) microarrays – microarranjo de DNA complementar* (baseado em hibridização (SAHA *et al.*, 1995); *Expressed Sequence Tag* (EST- Etiqueta de Sequência Expressa) (ADAMS *et al.*, 1991); *Serial Analysis of Gene Expression* (SAGE – Análise serial da expressão gênica) (VALCULESCU *et al.*, 1995) e suas versões aprimoradas – *Long Serial Analysis of Gene Expression* (LongSAGE – Análise serial longa da expressão gênica) (SAHA *et al.*, 2002) e *Super Serial Analysis of Gene Expression* (SuperSAGE – Análise super serial da expressão gênica) (MATSUMURA *et al.*, 2003); RNA-seq (MORTAZAVI, 2008) e mais recentemente a *Massive analysis of cDNA ends* (MACE – Análise massiva das extremidades de DNA complementar), baseados em sequenciamento (KAHL *et al.*, 2012).

As metodologias de transcriptômica mais antigas têm baixo poder de resolução, devido principalmente à (1) semelhança entre transcritos diferentes; (2) expressão de genes sobrepostos; (3) presença de transcritos alternativos. Para evitar o processamento errôneo dos dados, recomenda-se utilizar mais de uma metodologia de análise. Dessa forma, analisando os dados de diferentes metodologias é possível obter o máximo de informações a respeito do transcriptoma a ser analisado (NETO *et al.*, 2018)

A SuperSAGE , gera um abrangente perfil de transcrição, a partir de análises qualitativas e quantitativas de padrões de transcrição de milhões de *tags* de 26 pares de bases (pb) sequenciadas em um mesmo experimento, oriundas de um ou mais

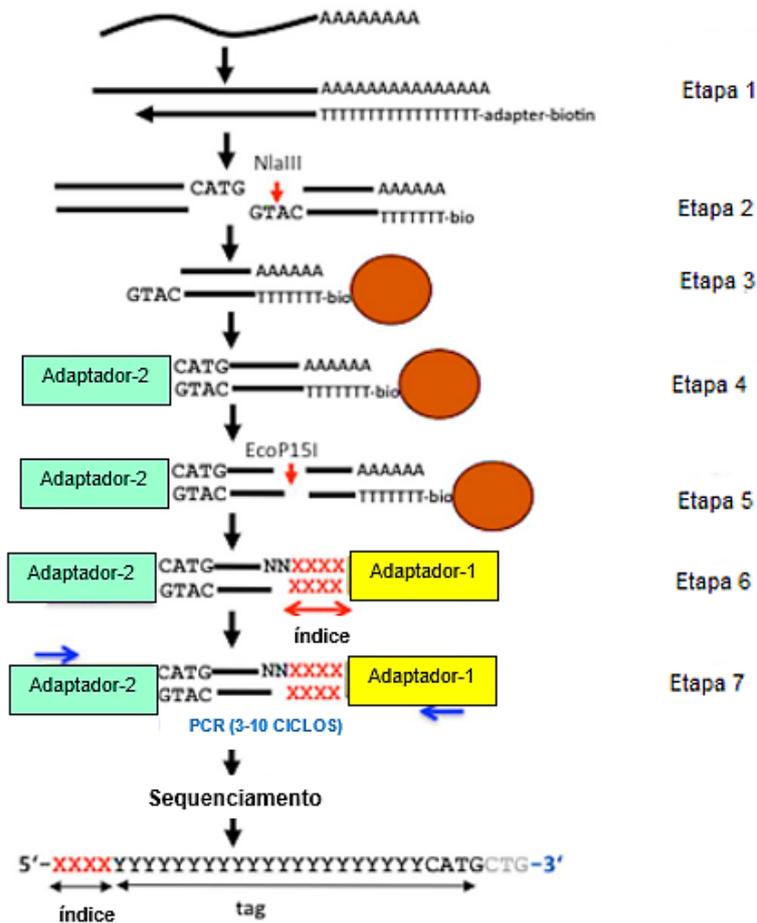
genomas de eucariotos por ensaio; baseando-se nas mudanças de frequência em RNA mensageiros possuidores de cauda poliA (MATSUMURA *et al.*, 2003). A técnica se torna ainda mais valiosa quando combinada com tecnologias de sequenciamento (de nucleotídeos) de alto desempenho (MATSUMURA *et al.*, 2010). Essa combinação, denominada *High-Throughput-SuperSAGE* ou *deepSuperSAGE*, fornece muito mais dados (três a quatro ordens de magnitude), com custos relativamente baixos, redução de tempo e esforços (MATSUMURA *et al.*, 2010).

HT-SuperSAGE mimetiza a técnica original, SuperSAGE (MATSUMURA *et al.*, 2003), até a etapa em que a *tags* de 26 pb são cortadas a partir de cDNAs de cadeia dupla. Porém, após esta etapa não ocorre a formação de “*ditags*” (duas *tags* com orientação invertida) como descrito pelo protocolo original, em vez disso, cada extremidade de uma *tag* simples é ligada a dois adaptadores distintos e os fragmentos “*tag* adaptados” são amplificados por PCR (*Polymerase Chain Reaction*) com número limitado de ciclos (MATSUMURA *et al.*, 2010).

A metodologia se inicia com a síntese de cDNA utilizando *primers* oligodT contendo o sítio de reconhecimento para EcoP15I. Posteriormente, esse cDNA é clivado pela enzima de ancoragem *NlaIII* que reconhece sítios 5'-CATG-3'. Após a primeira digestão com *NlaIII*, é adicionado o adaptador-2 e então prossegue-se uma segunda digestão, dessa vez com a enzima EcoP15I (enzima de restrição que cliva a uma distância de 26-27 pb de seu sítio de reconhecimento, denominada enzima de “*tagging*”). Em seguida, ocorre a ligação do adaptador-1, o qual, possui uma sequência índice de 4 pb para identificação da biblioteca. Os fragmentos seguem para PCR e posterior sequenciamento (Figura 9).

A geração de uma *tag* maior, com 26 pb, melhora consideravelmente o processo de anotação (associação das *tags* com seus genes correspondentes), permitindo uma anotação quase que perfeita através da ferramenta BLAST contra bancos de dados, nos quais genomas e coleções de cDNA estiverem disponíveis. A vantagem fica ainda mais evidente quando se compara com as outras metodologias anteriores a esta, como a SAGE, *tag* com 15 pb (VALCULESCU *et al.*, 1995) e a Long-SAGE, *tag* com 22 pb (MATSUMURA *et al.*, 2008), nas quais, a anotação se fez um problema inerente.

Figura 9 - Fluxograma da técnica HT-SuperSAGE. Traduzido de MATSUMURA *et al.*, (2010).



Finalmente, uma parte significativa dos estudos em transcriptômica é acompanhada por análises de bioinformática. Os desafios destas análises são quase tão grandes como os de procedimento experimental, tais como purificação do RNA, construção das bibliotecas de cDNA e o sequenciamento propriamente dito. Isto se deve a elevada quantidade de dados a serem analisados, bem como, os erros que podem ser introduzidos pelo sequenciamento e da complexidade inerente dos transcriptomas (DONG e CHEN, 2013).

**2.4.1 Ferramentas de bioinformática aplicadas ao estudo de genes estresse-responsivos em plantas**

Os mecanismos moleculares que compreendem a regulação da expressão são complexos e não são esclarecidos apenas pelo sequenciamento do genoma de um

determinado organismo. Esta etapa, denominada de genômica estrutural, é apenas o início de um processo que visa compreender como as estruturas do genoma funcionam. A próxima etapa, conhecida como genômica funcional, abrange análises que vão utilizar as informações oriundas da genômica estrutural para caracterização das funções gênicas (HIETER e BOGUSKI, 1997).

#### 2.4.1.1 Bancos de dados biológicos

Os bancos de dados biológicos constituem ponto central na bioinformática, permitindo que cientistas acessem uma ampla gama de dados de uma variedade de organismos. Estes repositórios podem ser divididos de acordo com os dados que armazenam, sendo os que abrangem todas as sequências de nucleotídeos e proteínas designados como bancos primários e aqueles de recurso mais especializado como bancos de identificação e estrutura de proteínas e específicos para determinados organismos (WHITFIELD *et al.*, 2005). Alguns exemplos de bancos de dados estão descritos na tabela 1.

Tabela 1 - Alguns exemplos de bancos de dados utilizados em transcriptômica vegetal

Tipo de banco	Descrição	Referência
Nucleotídico	RefSeq	<a href="http://www.ncbi.nlm.nih.gov/refseq/">www.ncbi.nlm.nih.gov/refseq/</a>
Nucleotídico	<i>Ensembl</i>	CUNNINGHAM <i>et al.</i> , (2018)
Proteico	<i>Entrez protein.</i>	<a href="http://www.ncbi.nlm.nih.gov/protein/">www.ncbi.nlm.nih.gov/protein/</a>
Proteico	UniProt ( <i>Universal Protein Resource</i> )	UNIPROT CONSORTIUM, (2019)
Especializado	Grassius ( <i>Grass Regulatory Information Server</i> )	YILLMAZ <i>et al.</i> , (2009)
Especializado	TAIR ( <i>The Arabidopsis Information Resource</i> )	POOLE, (2005)
Especializado	iTAK ( <i>Plant Transcription factor &amp; Protein Kinase Identifier and Classifier</i> )	ZHENG <i>et al.</i> , (2016)
Especializado	KEGG ( <i>Kyoto Encyclopedia of Genes and Genomes</i> )	KANEHISA, (2002)

#### 2.4.1.2 Alinhamento de sequências e montagem de transcriptomas

A etapa inicial para o entendimento das características de uma sequência desconhecida é o alinhamento de sequências. Basicamente, a sequência desconhecida é alinhada com sequências conhecidas disponíveis em bancos de dados, com o intuito de identificar resíduos comuns que são conservados ao longo da evolução (THOMPSON *et al.*, 2001). Existem dois métodos de alinhamento, o global e o local. No alinhamento global a semelhança é pontuada ao longo de todo o comprimento das sequências e o alinhamento local que identifica regiões similares entre as sequências, como domínios funcionais (NEEDLEMAN e WUNSCH 1970; SMITH e WATERMAN 1981).

Igualmente, existem dois tipos de alinhamento: o alinhamento par a par, que considera duas sequências, e o múltiplo que considera mais que duas sequências. As sequências podem ser nucleotídicas (DNA ou RNA) ou de aminoácidos (proteínas), no alinhamento podem ser introduzidos um ou mais *gaps* que indicam ganho ou perda de um resíduo, e evolutivamente equivalem a inserções ou deleções (*indel*). A programação dinâmica garante alinhamento ótimo entre duas sequências (alinhamento par a par), porém, é computacionalmente custoso. Dessa forma, foram desenvolvidos algoritmos heurísticos que provém alinhamentos adequados em menor tempo (WATERMAN, 1984).

O mais popular algoritmo heurístico de alinhamento é o BLAST (*Basic Local Alignment Search Tool*) (ALTSCHUL *et al.*, 1990). O método de alinhamento do BLAST é baseado em blocos, em que as sequências são “quebradas” em 11 e em três resíduos para nucleotídeos e proteínas, respectivamente. A noção de semelhança é incorporada no algoritmo através da utilização de uma matriz de *scores*, como por exemplo a BLOSUM62. Quando as palavras com *match* são encontradas, o BLAST estende a região através do alinhamento local (sem *gaps*) das duas extremidades, em seguida são identificados os elementos de maior *score* “HSP” (*High Score Pair*). Novas versões deste algoritmo foram desenvolvidas (ALTSCHUL *et al.*, 1997), alguns exemplos são o PSI-BLAST (*Position-Specific Iterated BLAST*) e o *Gapped BLAST*. O

PSI-BLAST procede de forma iterativa. Após alinhamento das regiões conservadas, o algoritmo constrói um consenso a partir das regiões alinhadas. Usando este consenso, procede então uma nova pesquisa sobre a base de dados. O processo finaliza quando nenhuma nova sequência é adicionada ao consenso. O *Gapped* BLAST é modificado para desencadear uma extensão com *gaps* para qualquer HSP que excede um *score* moderado.

Um exemplo de algoritmo heurístico para alinhamento múltiplo é o alinhamento progressivo (HOGEWEG e HESPER, 1984). No alinhamento progressivo, o alinhamento múltiplo é resolvido utilizando o alinhamento par-a-par indiretamente. Essa abordagem reúne todas as sequências progressivamente, onde o melhor alinhamento par-a-par é considerado primeiro. O alinhamento é resolvido pela utilização de uma árvore guia (FENG e DOOLITTLE, 1987), onde cada ramo representa uma sequência a ser alinhada, cada nó é associado com um alinhamento múltiplo das sequências em sua subárvore correspondente, dessa forma, o alinhamento múltiplo de todas as sequências consideradas é associado a raiz do nó. Este algoritmo é utilizado em programas como CLUSTAL (HIGGINS e SHARP, 1988), T-coffee (NOTREDAME *et al.*, 2000), MAP (HUANG, 1994), MULTALIGN (CORPET, 1988) dentre outros.

A identificação do conjunto completo de transcritos é base dos estudos em transcriptômica. As *reads* obtidas de plataformas de sequenciamento são pequenas sendo então necessário a reconstrução do tamanho total do transcrito, que constitui um desafio na área. Basicamente, após o sequenciamento, as *short-reads* são pré-processadas para remover erros de sequenciamento e quaisquer artefatos (ex.: adaptadores), em seguida, as *reads* são montadas afim de observar suas abundâncias ('análise da expressão') (MARTIN e WANG, 2011). Para auxiliar no processo de montagem, geralmente são preferidas *long reads* pois reduzem a complexidade da montagem. A escolha de protocolos *paired-end* (ambas as extremidades dos fragmentos de DNA são sequenciadas) também atenuam o problema das *short-reads*.

Existem duas estratégias para realizar a montagem de um transcriptoma: (1) baseada em referência ou montagem "*ab initio*" e (2) montagem *de novo*. Na estratégia por referência alguns problemas são contornados, como a contaminação ou artefatos

de sequência, estes não são esperados para alinhar com o genoma de referência. Porém também apresenta desvantagens, pois depende fortemente de um genoma de referência de boa qualidade, um problema inerente quando se trata de organismos não-modelo. Na montagem *de novo* não há necessidade de uma referência, podendo também cobrir transcritos de regiões genômicas que estão ausentes em montagens disponíveis, as principais dificuldades são o recurso computacional requerido e a necessidade de sequenciamento com maior profundidade (MARTIN *et al.*, 2010; MARTIN e WANG, 2011). Alguns exemplos de ferramentas, utilizadas na montagem de transcriptomas de plantas são: TopHat (TRAPNELL *et al.*, 2009) e Trinity (GRABHERR *et al.*, 2011).

#### 2.4.1.3 Anotação funcional

A atribuição de função biológica é a etapa conhecida como anotação funcional. Esse processo frequentemente depende da transferência de informação de um gene a outro, também denominado de anotação baseada em homologia. Depende que a informação a respeito dos produtos gênicos esteja disponível e assume que a similaridade é resultado de homologia funcional (BOLGER *et al.*, 2018). Existem várias ferramentas destinadas a este objetivo:

*Gene Ontology* (GO) – é a ferramenta mais utilizada para anotação funcional, que fornece ‘termos GO’ associados aos produtos gênicos de interesse divididos em três categorias: ‘*Biological Process*’ (descreve o gene em termos de eventos ou funções moleculares), ‘*Cellular Component*’ (descreve a localização de uma proteína), ‘*Molecular Function*’ (descreve as habilidades de um produto gênico em nível molecular) (GENE ONTOLOGY CONSORTIUM, 2015).

TransDecoder – busca por *frames* que codificam domínios funcionais dentro de sequências de transcritos (<http://transdecoder.sf.net>).

PFAM – é um excelente recurso que identifica domínios, representados como modelo oculto de Markov (*Hidden Markov Model*-HMM), dentro de uma sequência, e atualmente pode identificar >17.000 famílias proteicas (EL-GEBALI *et al.*, 2019).

BLAST2GO – é um *pipeline* automatizado que é utilizado para inferir anotações GO, que incorpora dados do InterProScan e KEGG (CONESA e GOTZ, 2008). No entanto, alguns dos recursos não são livres e requerem licença.

MEME *suite* – é um servidor *web* para descoberta e análise de motivos de sequência que representam importantes características como sítios de ligação de fatores de transcrição (BAILEY *et al.*, 2009).

MapMan BIN *ontology* – ferramenta desenvolvida para visualização de dados ômicos em vias de plantas e que também abrange em torno de 2.000 termos de ontologia (THIMM *et al.*, 2004).

Existem também ferramentas de recurso mais específico, são estes alguns exemplos:

SignalP – identifica a presença de peptídeo sinal em sequências de aminoácidos, que é essencial em proteínas secretadas. A última versão é capaz de distinguir entre o peptídeo sinal e hélices transmembrana (PETERSEN *et al.*, 2011).

psRNATarget – analisa miRNAs e seus alvos e possui duas importantes características: (i) *match* reverso complementar entre o miRNA e seu alvo (ii) avaliação da acessibilidade do sítio no alvo calculando a energia não emparelhada (UPE) necessária para "abrir" a estrutura secundária em torno do sítio alvo do miRNA (DAI e ZHAO, 2018).

Musite – ferramenta especificamente desenvolvida para prever sítios de fosforilação, importante evento regulatório para muitos aspectos celulares (GAO e XU, 2010).

#### 2.4.1.4 Desafios da bioinformática

O advento da bioinformática permitiu gerar enormes quantidades de dados, e a organização e armazenamento destes dados conduziu a importantes iniciativas nacionais e internacionais. Inicialmente, os bancos de dados eram arquivos simples e com o passar do tempo foram aumentando em tamanho e se relacionando com outros bancos de dados, para garantir uma melhor performance ao usuário. Um dos maiores desafios relacionados a organização e armazenamento é como integrar dados

genômicos com dados de mapeamento genômico, expressão gênica, função proteica, interação proteína-proteína dentre outros (MATHUR, 2018).

As técnicas de mineração de dados, utilizadas para analisar padrões em grandes conjuntos de dados e representa-los de forma humanamente legível, é frequentemente dificultada por características inerentes de bancos de dados biológicos como seu tamanho, número e diversidade. O nível de conhecimento entre os usuários também passou a ser uma dificuldade em prover mecanismos de acessos apropriados para todos (MATHUR, 2018). Algumas outras dificuldades associadas ao manejo dos dados são a falta de padronização nos formatos, falta de interfaces comuns aos dados, inconsistências em identificadores dentre outros. Os custos para o arquivamento de dados são de fato consideráveis, mesmo com métodos avançados de redução da informação, e a decisão de quais *data sets* devem arquivados ou armazenados tem sido um ponto de discussão (FULLER *et al.*, 2013).

Um outro desafio da bioinformática é que muitos programas são computacionalmente exaustivos, sendo crucial o desenvolvimento de novos algoritmos em termos de otimização de recursos. As áreas que a necessidade de novos algoritmos é constante são: alinhamento de sequências, predição de estrutura e função de proteínas e interação proteína-proteína (MATHUR, 2018). Uma questão que tem sido debatida dentro da bioinformática diz respeito a hipótese sendo dirigida pelos dados, ou seja, a coleção de dados vem antes do estabelecimento de uma hipótese testável, podendo constituir um viés na interpretação dos dados (FULLER *et al.*, 2013).

Finalmente, o rápido e constante crescimento da informação biológica disponível tornaram necessário o melhorar a habilidade de estudantes de graduação no campo da bioinformática, visto que tem sido expandido apenas entre pós-graduandos (TARCZY-HORNOCH e MINIE, 2005; MATHUR 2018). Um exemplo deste tipo de iniciativa é o EDUCOLLAB do NCBI que desenvolve vários programas de treinamentos introdutórios e avançados para o uso de ferramentas de bioinformática, que incluem análise de similaridade, expressão, estrutura de proteínas, análise filogenética dentre outros (GEER *et al.*, 2002; [Online] <http://www.ncbi.nlm.nih.gov/Class/NAWBIS/>)

### 3 MATERIAL E MÉTODOS

#### 3.1 MATERIAL VEGETAL, ENSAIO DE DÉFICIT HÍDRICO E BIBLIOTECAS HT-SuperSAGE

As *unitags* HT-SuperSAGE (26 pb) disponíveis para o presente estudo foram oriundas de ensaio de oito cultivares de cana-de-açúcar selecionados do programa de melhoramento genético do Centro de Tecnologia Canavieira (CTC, Piracicaba, Brasil), sendo quatro tolerantes (CTC6, CTC15, SP83-2847, e SP83-5073) e quatro sensíveis à seca (CTC9, CTC13, SP90-3414, e SP90-1638), de acordo com observações em campo e análises fisiológicas, incluindo conteúdo relativo de água e teor de clorofila (THIEBAUT *et al.*, 2012; THIEBAUT *et al.*, 2014). Parte das plantas de cada grupo, cultivadas em potes de 40 L, em casa de vegetação [temperaturas de  $30,2 \pm 5,7$  °C (máxima),  $16,8 \pm 1,9$  °C (média) e  $9,3 \pm 3,0$  °C (mínima), humidade relativa [ $71,5 \pm 5,1\%$  (média)], com fotoperíodo natural, foram submetidas a 24 h de supressão de rega, conforme descritos por Kido *et al.* (2012) e da Silva *et al.* (2017). Resumidamente, plantas em desenho experimental aleatório (seis repetições), com irrigação diária ( $4 \text{ L h}^{-1}$ ) até os três meses de idade, tiveram a irrigação interrompida por 24 horas, quando raízes foram coletadas, congeladas em nitrogênio líquido, e armazenadas ( $-80^{\circ}\text{C}$ ) até extração do RNA total, com uso de Trizol® (Invitrogen). As amostras de RNAs, após tratamento com DNase (Qiagen®) e purificação com RNeasy Mini kit (Qiagen) tiveram quantidades similares de cada cultivar misturadas conforme o grupo, para gerarem as bibliotecas [tolerante estressado (SD24T), tolerante controle negativo (SDTC), sensível estressado (SD24S) e sensível controle negativo (SDSC)] HT-SuperSAGE (MATSUMURA *et al.*, 2003). Após sequenciamento por *Illumina Genome Analyzer II* (KIDO *et al.*, 2012), as tags de 26 pb sequenciadas apenas uma vez (*singlets*) foram eliminadas. Das *unitags* (tags diferentes) observadas, aquelas diferencialmente expressas ( $p < 0,05$ ) foram identificadas a partir da distribuição de Poisson (AUDIC e CLAVERIE, 1997) implementada no software DiscoverySpace (v.4.01) (ROBERTSON *et al.*, 2007). A modulação da expressão gênica de cada *unitag* (FC: *Fold Change*) foi calculada como a razão das frequências observadas nas bibliotecas estressada em relação ao controle negativo.

### 3.2 ANOTAÇÃO DE TRANSCRITOS CoREG, IDENTIFICAÇÃO DE DOMÍNIOS FUNCIONAIS E ANÁLISE *IN SILICO* DA EXPRESSÃO

Sequências representativas de CoREGs de 24 famílias (incluindo o grupo “Others”) foram obtidas da base de dados iTAK (<http://itak.feilab.net/cgi-bin/itak/index.cgi>) (ZHENG *et al.*, 2016). ESTs de cana-de-açúcar (*Saccharum spp.*), disponíveis nos bancos públicos do NCBI (<https://www.ncbi.nlm.nih.gov/>), foram alinhadas via BLASTn (*e-value cutoff*  $e^{-20}$ ) contra as sequências CoREGs. Anotações das ESTs foram confirmadas em alinhamentos BLASTn ( $e^{-20}$ ) contra transcritos RefSeq de monocotiledôneas (555.398) (<https://www.ncbi.nlm.nih.gov/refseq/>). Sequências com anotações não correspondentes a CoREGs foram descartadas das análises. As ESTs anotadas como CoREGs, após tradução pela ferramenta TransDecoder (v 5.0.2; <http://transdecoder.github.io>), tiveram potenciais domínios identificados através da ferramenta NCBI *Batch Web Conserved Domain Search* (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>; MARCHLER-BAUER *et al.*, 2015).

*Unitags* observadas na resposta tolerante (SD24T vs. SDTC) e na sensível (SD24S vs. SDSC) foram alinhadas via BLASTn ( $e^{-20}$ ) contra as ESTs de cana-de-açúcar identificadas como potenciais CoREGs. Alinhamentos *unitags*-ESTs aceitáveis apresentaram *score* > 44 (score de 52 correspondeu a 100% de identidade), máximo de um *mismatch* em 26 pb, orientação plus/plus, e extremidade 5'-CATG da *unitag* preservada. A clusterização hierárquica de *unitags* diferencialmente expressas (*p-value* < 0,0001) foi realizada pelo software Cluster 3.0 (v.1.52) (EISEN *et al.*, 1999), utilizando parâmetros padrões e valores de FC das *unitags*, cujos dendrogramas foram visualizados pelo software TreeView (PAGE, 1996). Diagramas de Venn foram gerados a partir da ferramenta InteractiVenn (<http://www.interactivenn.net/index.html>) (HEBERLE *et al.*, 2015).

### 3.3 ANÁLISE DE TERMOS GO E DE FTs

As ESTs contendo *unitags* diferencialmente expressas ( $p < 0,05$ ) nos tratamentos dos *bulks* de cultivares foram alinhadas via BLASTn (*e-value cutoff*  $e^{-20}$ )

contra sequências CDS de *S. bicolor* disponíveis na base de dados PlantRegMap (<http://plantregmap.cbi.pku.edu.cn/index.php>) (TIAN *et al.*, 2019). Essas sequências representando as *unitags* diferencialmente expressas foram submetidas a análise de enriquecimento de termos de ontologia gênica (GO) e de FTs (*threshold p-value*  $\leq 0,01$ ), ambas a partir de ferramentas do próprio site.

### 3.4 MAPEAMENTO *IN SILICO* DE *LOCI* CoREGs EM GENOMAS DE GRAMÍNEAS

ESTs de CoREGs com domínios confirmados foram submetidas ao BLASTn (*e-value cutoff* e-20), contra transcritos de sorgo (*Sorghum bicolor*, Sbicolor\_313\_v3.1.transcript), milho (*Zea mays*, Zmays\_284\_5b+transcript) e arroz (*Oryza sativa*, Osativa\_323\_v7.0.transcript), visando identificar *loci* CoREGs em genomas de gramíneas taxonomicamente relacionadas. Sequências ortólogas para essas gramíneas foram obtidas a partir dos *best hits* recíprocos (*Bidirecional Best-Hit*, BBH), via BLASTp (*e-value cutoff* e-10) das ESTs com domínios confirmados contra os proteomas destas espécies. A partir do BLASTx (*e-value cutoff* e-10), entre os CoREGs ortólogos de cana-de-açúcar com sorgo contra os proteomas de milho e arroz, foram construídos ideogramas circulares com a ferramenta Circos 0.64 (KRZYWINSKI *et al.*, 2009), facilitando visualizar possíveis duplicações gênicas, sintenias e colinearidades de genes ligados, nas espécies analisadas.

### 3.5 EXTRAÇÃO DE RNAs E SÍNTESE DE cDNAs

RNAs das cultivares foram extraídos e purificados conforme descritos em KIDO *et al.*, (2012). A integridade dos RNAs foi avaliada em gel de agarose 1,5% corado com Blue-Green Loading Dye (LGC Biotecnologia®), e a quantificação feitas utilizando-se o Quant-iT™ RNA Assay Kit (Invitrogen®), em fluorômetro Qubit (Invitrogen®). RNAs totais (1 µg) foram convertidos em cDNAs utilizando-se o kit ImProm-II™ Reverse Transcription System (Promega®), sendo estes ressuspensos em 20 µL. As sínteses de cDNA incluíram uma incubação em termociclador a 42°C for 2 min. em gDNA Wipeout Buffer (para eliminar DNA genômico), 42°C por 15 min. (transcrição reversa) e 95°C for 3 min. (inativação da enzima). As amostras de cDNAs foram armazenadas em -20°C até as reações de qPCR, cujos os ensaios seguiram as

diretrizes do protocolo MIQE (*The Minimum Information for Publication of Quantitative Real-Time PCR Experiments*; BUSTIN *et al.*, 2009).

### 3.6 DESENHO DE *PRIMERS*, EFICIÊNCIA DE AMPLIFICAÇÃO E ANÁLISES DE qPCR

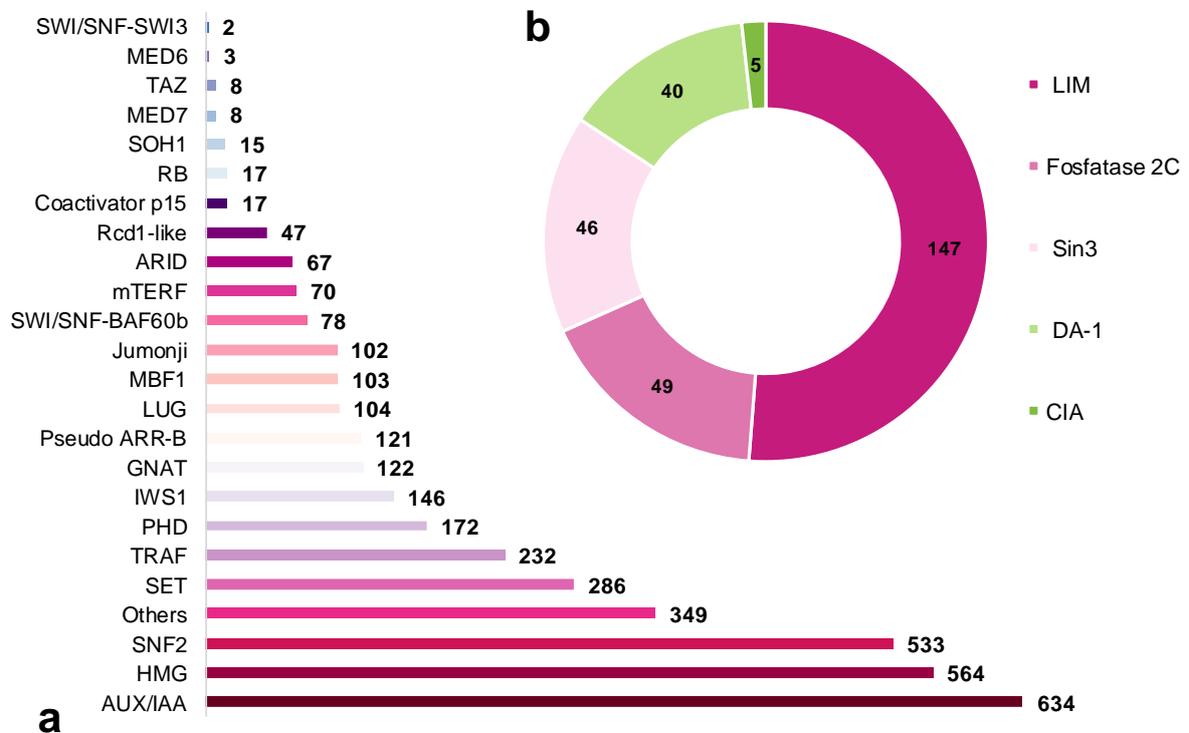
*Primers* para qPCR foram desenhados a partir das ESTs de CoREGs de cana-de-açúcar contendo *unitags* DE, com ajuda da ferramenta Primer3 (ROZEN e SKALETSKY, 2000), ajustando-se os parâmetros de: tamanho do *amplicon* (70-200 bp), conteúdo GC (45-55%), temperatura de dissociação (50-80°C), e CG clamp (um). Os alvos selecionados para *primers* foram TRAF (BTB/POZ and MATH domain-containing protein 1), Pseudo-ARR (*two-component response regulator* ORR9), Rcd1-like (CCR4-NOT *transcription complex subunit* 9), IWS1 (*probable mediator of RNA polymerase II transcription subunit* 26c) e LUH (*transcriptional corepressor* LEUNIG\_HOMOLOG-like). As especificidades dos *amplicons* foram analisadas em curvas de dissociação geradas entre 65-95°C (20 min.), após 40 ciclos de qPCR. A eficiência de amplificação ( $E = 10^{-1/\text{standard curve slope}}$ ; RASMUSSEN, 2001) de cada alvo foi derivada de curvas de calibração feitas com cDNAs das cultivares, em diluições seriadas (sem diluição, 10<sup>-1</sup>, 10<sup>-2</sup>, 10<sup>-3</sup> e 10<sup>-4</sup>). As reações de qPCR foram feitas em termociclador LineGene 9660 (Bioer®), utilizando-se SYBR *Green* como sistema de detecção, e volume final de 10 µL, sendo um µL de cDNA da amostra (10<sup>-1</sup>), cinco µL de GoTaq® qPCR Master Mix (Promega), dois µL de água ultra-pura, e um µL de cada primer (0.05 µM). As reações de RT-qPCR foram feitas em triplicata técnica e biológica, controle negativo (sem cDNA) e dois genes de referência (GAPDH e αTubulina) devidamente testados para o fim. As reações seguiram a seguinte programação: desnaturação inicial a 95°C (2 min.), seguido por 40 ciclos de 95°C (15 s) e 60°C (60 s). Ciclos de quantificação (Cq), temperatura de dissociação, e quantificação absoluta e relativa, foram estimados com o software que acompanha o termociclador. Dados de expressão foram calculados com o software REST (versão 2.0.13) (PFAFFL *et al.*, 2002).

## 4 RESULTADOS

### 4.1 ESTs CoREG E DISTRIBUIÇÃO EM FAMÍLIAS

A análise BLASTn (e-value  $e^{-20}$ ) com ESTs de cana-de-açúcar (285.216 transcritos, NCBI) e sequências (nt/aa) CoREGs da base de dados iTAK (2.067 sequências de *S. bicolor*, *Z. mays* e *O. sativa*) resultou em 25.146 alinhamentos aceitáveis, envolvendo 4.980 ESTs de cana-de-açúcar e 1.585 sequências CoREGs, representantes das 24 famílias disponíveis, inclusive o grupo *Others*. As ESTs similares às sequências CoREGs tiveram as anotações comparadas com as de transcritos RefSeq, em alinhamentos BLASTn (384.975; e-value  $e^{-20}$ ), que envolveram 4.980 ESTs e 25.286 transcritos (de um total de 55.398). A curagem manual mostrou que 3.800 ESTs corresponderam a 507 potenciais CoREGs. Nas ESTs curadas, todas as 24 famílias estiveram representadas, sendo as abundâncias ilustradas na Figura 10a. ESTs com anotações divergentes (1.180) foram excluídas das análises (Figura 10b). Compondo o grupo *Others* observou-se cinco anotações diferentes, sendo as mais representativas: *LIM domain* (147 ESTs), *phosphatase 2C* (49 ESTs), *Sin3* (46 ESTs), *DA1-protein* (40 ESTs) e *CIA (Chloroplast Import Apparatus*; cinco ESTs) (Figura 10c). Dentre as 3.800 ESTs codificadoras de CoREGs foram identificadas 449 ESTs de anotação “não caracterizada”, mas que foram consideradas nas demais análises.

Figura 10 - (a) Distribuição das ESTs de cana-de-açúcar nas famílias de correguladores oriundas do iTAK (<http://itak.feilab.net/cgi-bin/itak/index.cgi>) com anotações confirmadas pelo RefSeq mrna. (c) As anotações referentes ao grupo de CoREGs não caracterizado “Others”.



#### 4.2 CoREGs ASSOCIADOS A *unitags* HT-SuperSAGE

Das *unitags* das quatro bibliotecas HT-SuperSAGE disponíveis (205.975 *unitags* de 26 pb; KIDO *et al.*, 2012), alinhamentos aceitáveis *unitag*-EST envolveram 746 *unitags* em 390 ESTs CoREGs curadas de cana-de-açúcar. Desse montante, 107 *unitags* (74 CoREGs associados) foram declaradas diferencialmente expressas (DE;  $p$ -value < 0.05) após 24 h de supressão de rega. Na resposta tolerante (contraste TC; SD24T vs SDTC), 72 *unitags* foram induzidas (UR; *up regulated*) e 35 reprimidas (DR; *down regulated*). Na resposta sensível (contraste SC; SD24S vs SDSC), das 113 *unitags* DE (86 CoREGs), 61 *unitags* foram UR e 52 DR (Tabela 2).

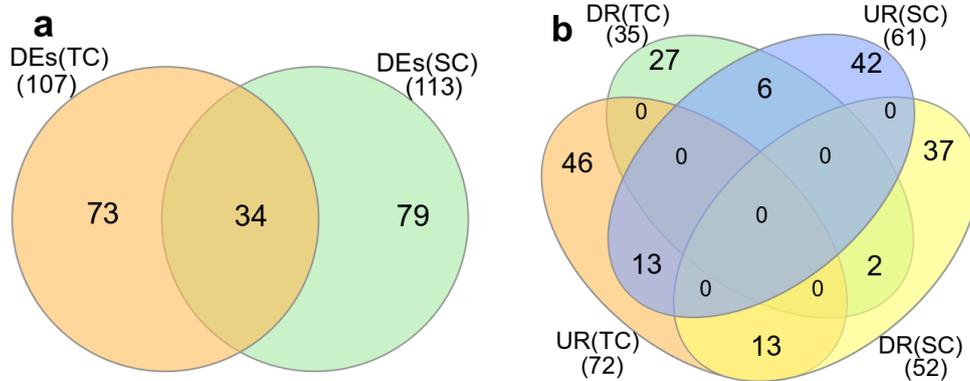
Tabela 2 - Total de *unitags* HT-SuperSAGE diferencialmente expressas (DE; *p-value* < 0.05), associadas a ESTs CoREGs, em raízes de cana-de-açúcar após 24 h de supressão de rega, quando comparadas a expressão dos respectivos controles irrigados.

<b>Contraste</b>	<b>DE</b>	<b>UR</b>	<b>DR</b>
Tolerante (SD24T vs SDTC)	107	72	35
Sensível (SD24S vs SDSC)	113	61	52

Bibliotecas: SDT24 (*bulk* dos quatro acessos tolerantes sob estresse); SDTC (*bulk* dos quatro acessos tolerantes sem estresse); SDS24 (*bulk* dos quatro acessos sensíveis sob estresse); SDSC (*bulk* dos quatro acessos sensíveis sem estresse).

Das *unitags* DE associadas a ESTs CoREGs, 73 foram exclusivas da resposta tolerante e 79 da sensível (Figura 11a). Destas 73 *unitags*, 21 (9 UR, 12 DR) não foram detectadas na resposta sensível. Das detectadas, 52 (37 UR, 15 DR) não modularam expressivamente a expressão. Das 79 *unitags* DEs exclusivas de SC, 14 (8 UR; 6 DR) não foram detectadas na resposta tolerante, e das detectadas, 65 (34 UR; 31 DR) não modularam diferencialmente. Ainda, as duas respostas compartilharam 34 *unitags* DE, algumas de regulação concordante (13 UR; 02 DR) e algumas divergentes (13 induzidas para TC e reprimidas para SC; seis reprimidas para TC e induzidas para SC). *Unitags* associadas a CoREGs e exclusivas da resposta tolerante foram 46 UR e 27 DR. Aquelas exclusivas da resposta sensível foram 42 UR e 37 DR. Concordantes em ambas as respostas foram 13 *unitags* induzidas e duas reprimidas. Adicionalmente, 13 *unitags* induzidas na resposta tolerante foram reprimidas na resposta sensível, enquanto que seis *unitags* reprimidas na resposta tolerante foram induzidas na sensível (Figura 11b).

Figura 11 - Diagramas de Venn ilustrando *unitags* exclusivas e compartilhadas entre os dois *bulks* de acessos (SD24TvsC[TC] e SD24SvsC[SC]) entre aquelas diferencialmente expressas (DEs  $p$ -value < 0,05) e aquelas induzidas (UR) e reprimidas (DR).



As *unitags* associadas a CoREGs e identificadas DE na resposta tolerante representaram 18 das 24 famílias investigadas. As mais abundantes foram AUX/IAA (33 *unitags*), SNF2 (17 *unitags*) e HMG (16 *unitags*) (Figura 12a). A resposta sensível envolveu 20 famílias, sendo Aux/IAA (24 *unitags*), SNF2 (15 *unitags*) e HMG (16 *unitags*) também as mais abundantes (Figura 12b). Por sua vez, a família RB (SD170902) foi associada apenas a resposta tolerante, enquanto que as famílias Jumonji (SD265694; SD11897; SD252556), mTERF (SD127604; SD74486) e SOH1 (SD190126) foram exclusivas da resposta sensível. Representante das famílias MED7, TAZ e SWI-SNF/SWI3 não foram associadas às *unitags* DE de nenhuma das duas respostas (Figura 12). Por outro lado, considerando as *unitags* UR nas duas respostas, 16 famílias foram observadas na resposta tolerante e 13 na sensível, sendo Pseudo-ARR (SD181421; SD6438), RB (SD133947), Rcd1-like (SD194342; SD262454) e SWI-SNFBAF60 (SD171600; SD171582) exclusivas da resposta tolerante. Curiosamente, apesar da resposta sensível ter apresentado menos *unitags* induzidas (Tabela 2), estas abrangeram mais famílias CoREG que na resposta tolerante, na qual houve uma certa concentração de *unitags* UR nas famílias AUX/IAA (28 *unitags*) e HMG (12 *unitags*). O número de famílias e respectivos números de *unitags* (UR; DR; n.s.; n.d.) nas duas respostas estão descritos na tabela 3.

Figura 12 - Distribuição das *unitags* HT-SuperSAGE nas famílias CoREG identificadas nas ESTs de cana-de-açúcar. Em (a) *unitags* DE ( $p < 0,05$ ); em (b) apenas as *unitags* UR.

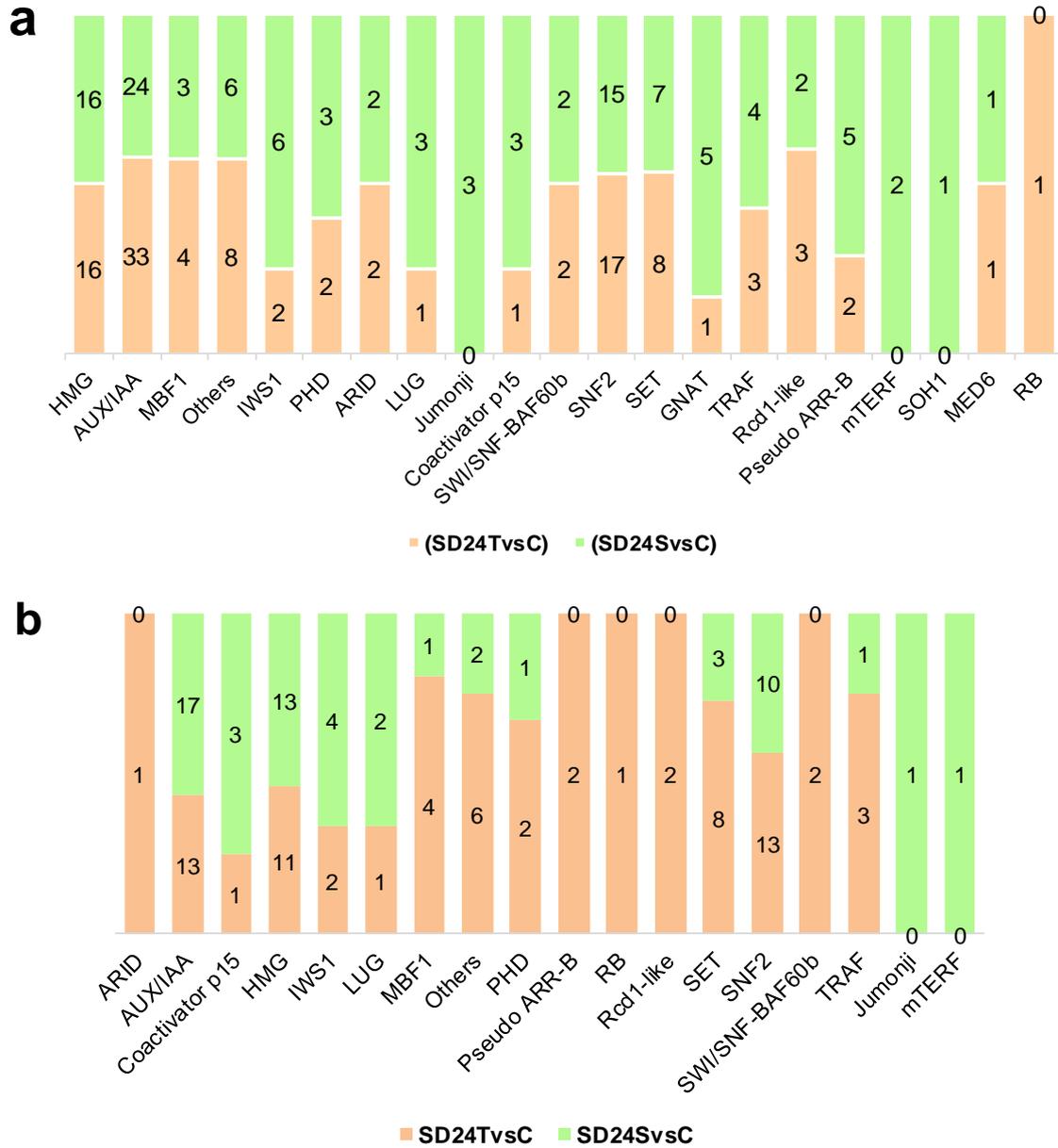


Tabela 3 - Número de *unitags* associadas a famílias de corre reguladores em biblioteca HT-SuperSAGE de raízes de acessos contrastantes (tolerante [SD24TvsSDTC]; sensível [SD24SvsSDSC]) de cana-de-açúcar após 24 h de supressão rega. As *unitags* são representadas com base na regulação detectada (UR: *up regulated*; DR: *down regulated*; n.s.: não significativo; n.d.: não detectada)

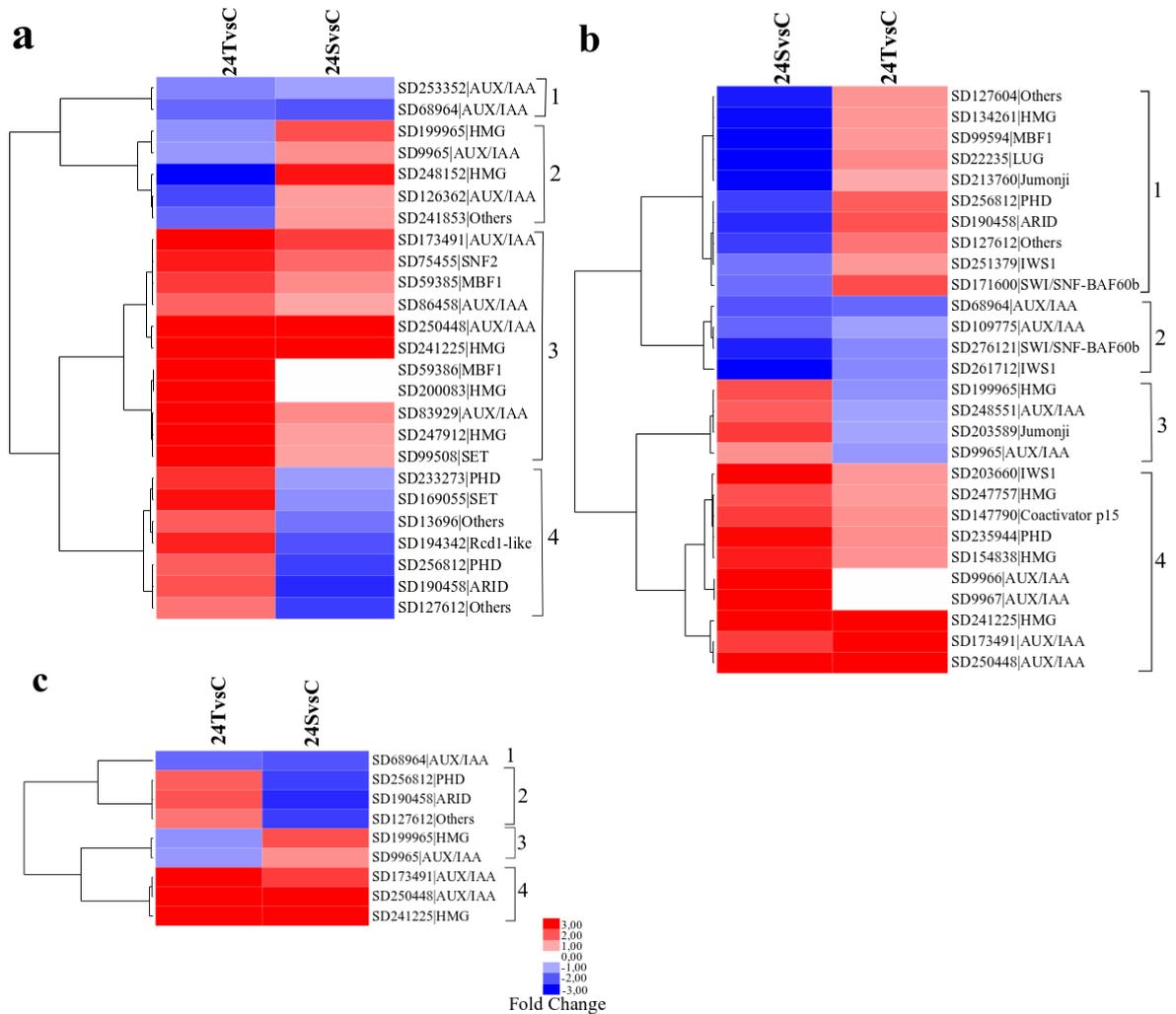
Família CoREG	SD24TvsSDTC [UR]	SD24SvsSDSC [UR]	SD24TvsSDTC [DR]	SD24SvsSDSC [DR]	SD24TvsSDTC [n.s.]	SD24SvsSDSC [n.s.]	SD24TvsSDTC [n.d.]	SD24SvsSDSC [n.d.]
ARID	1	0	1	2	9	5	1	5
AUX/IAA	13	17	20	7	86	75	25	45
Coactivator p15	1	3	0	0	8	5	4	5
GNAT	0	2	1	3	27	21	3	5
HMG	11	13	5	3	27	24	12	15
IWS1	2	4	0	2	29	21	9	13
Jumonji	0	1	0	2	14	10	3	4
LUG	1	2	0	1	12	7	2	5
MBF1	4	1	0	2	15	5	1	12
MED6	0	1	1	0	2	1	0	1
MED7	0	0	0	0	1	3	3	1
mTERF	0	1	0	1	9	7	0	0
Others	6	2	2	4	24	24	8	10
PHD	2	1	0	2	32	30	13	14
Pseudo ARR-B	1	0	0	5	13	8	4	5
RB	1	0	0	0	3	5	3	2
Rcd1-like	2	0	1	2	7	9	4	3
SET	8	3	0	4	45	40	11	17
SNF2	13	10	4	5	84	66	19	39
SOH1	0	0	0	1	2	0	0	1
SWI/SNF-BAF60b	2	0	0	2	14	15	5	4
TAZ	0	0	0	0	1	2	1	0
TRAF	3	1	0	3	30	32	14	11

Com o objetivo de contrastar as respostas dos dois grupos de cultivares, as *unitags* DE de cada resposta (contraste TC e SC), tendo se aplicado limiar mais restrigente ( $p\text{-value} < 1e^{-4}$ ), bem como aquelas em comum aos dois *bulks*, foram comparadas por análise de clusterização hierárquica, a partir dos valores FC para modulação da expressão. Das *unitags* DE da resposta tolerante, comparadas com respectivos valores da resposta sensível foi possível observar quatro grupos (Figura 13a), compostos por: (1) dois CoREGs DR da resposta TC, que na resposta SC foram DR (SD68964) e n.s. (SD253352); (2) cinco CoREGs DR na resposta TC, que na resposta SC foram UR (SD199965; SD9965; SD248152) e n.s. (SD126362; SD241853); (3) 11 CoREGs UR na resposta TC, que na resposta SC foram UR (SD173491; SD75455; SD59385; SD250448; SD241225), n.s. (SD86458; SD83929; SD247912; SD99508) e n.d. (SD59386; SD200083), (4) sete CoREGs UR na resposta TC, que na resposta SC foram DR (SD13696; SD194342; SD256812; SD190458; SD127612) e n.s. (SD233273; SD169055).

Das *unitags* DE da resposta SC, que quando comparadas com a modulação da resposta TC, permitiu observar quatro grupos (Figura 13b) compostos por: (1) 10 CoREGs DR na resposta SC, que na resposta TC foram UR (SD127604; SD99594; SD256812; SD190458; SD127612; SD171600) e n.s. (SD134261; SD22235; SD213760; SD251379); (2) quatro CoREGs DR na resposta SC, que na resposta TC foram DR (SD68964) e n.s. (SD109775; SD276121; SD261712); (3) quatro CoREGs UR na resposta SC, que na resposta TC foram DR (SD199965; SD9965) e n.s. (SD248551; SD203589); (4) 10 CoREGs UR na resposta SC, que na resposta TC foram UR (SD247757; SD241225; SD173491; SD250448), n.s. (SD203660; SD147790; SD235944; SD154838) e n.d. (SD9966; SD9967).

Das *unitags* DE da resposta SC, comparadas com sua respectiva resposta DE de TC permitiu observar quatro grupos (Figura 13c) formados por: (1) CoREG com *unitag* DR (SD68964) em ambos os contrastes; (2) três CoREGs com *unitags* (SD256812; SD190458; SD127612) UR em TC e DR em SC; (3) dois CoREGs com *unitags* (SD199965; SD9965) DR em TC e UR em SC; (4) três CoREGs com *unitags* (SD173491; SD250448; SD241225) UR em ambos os contrastes

Figura 13 - Clusterização hierárquica feita a partir dos dados de modulação da expressão (FC; *Fold Change*) nos dois *bulks* de cultivares, de parte das *unitags* DE (diferencialmente expressas,  $p$ -value < 0,0001), de acordo com as comparações: DE tolerante (SD24T vsC) e respectiva modulação no sensível (SD24SvsC) (a); DE sensível e respectiva modulação no tolerante (SD24T vsC) (b); DE em comum aos dois *bulks* de cultivares (c).



### 4.3 ENRIQUECIMENTO DE TERMOS GO

Das ESTs CoREGs (74) relacionadas às *unitags* DE ( $p < 0,05$ ) da resposta tolerante (106), em análise BLASTn (*e-value*  $e^{-20}$ ) contra sequências CDS de *S. bicolor* (33.971) da base de dados PlantRegMap, observou-se que as 74 ESTs foram similares a 59 sequências CDS de sorgo, das quais 53 apresentavam anotação GO. Igualmente, em análise BLASTn (*e-value*  $e^{-20}$ ) contra as sequências CDS de sorgo, das 86 ESTs CoREGs relacionadas às 113 *unitags* DE ( $p < 0,05$ ) da resposta sensível, estes foram similares a 74 sequências CDS de sorgo, sendo 70 com anotação GO. Termos GO enriquecidos associados às CDS/ESTs com *unitags* DE da resposta tolerante foram 204, com destaque nas categorias de Processo Biológico para a “via de sinalização ativada por auxina” (GO:0009734), “resposta celular ao estímulo da auxina” (GO:0071365) e “via de sinalização mediada por hormônios” (GO:0009755). Nas categorias de Função Molecular, os termos enriquecidos abrangeram “ligação a proteína” (GO:0005515), “atividade de histona lisina N-metiltransferase” (GO:0018024) e “ligação” (GO:0005488). Para a grande categoria Componente Celular, o destaque incluiu “núcleo” (GO:0005634), “organela intracelular” (GO:0043229) e “organela” (GO:0043226). Por sua vez, na resposta sensível, os termos GO enriquecidos, associados às 86 *unitags* DE ( $p < 0,05$ ), foram 231. Nas categorias de Processo Biológico, “regulação da expressão gênica” (GO:0010468), “regulação do processo metabólico de macromoléculas” (GO:0060255), e “regulação de processos metabólicos” (GO:0019222) foram os destaques. Em Função Molecular, os termos mais enriquecidos foram “ligação a proteína” (GO:0005515), “atividade de coativador da transcrição” (GO:0003713) e “atividade de cofator da transcrição” (GO:0003712), enquanto que para Componente Celular os termos mais enriquecidos foram “núcleo” (GO:0005634), “intracelular” (GO:0005622) e “organela intracelular” (GO:0043229). Ainda, um total de 150 termos GO enriquecidos foram observados compartilhados entre as duas respostas. Entretanto, 54 termos GO foram exclusivos da resposta tolerante, enquanto que 81 termos foram exclusivos da resposta sensível (Figura 14a). Dos termos GO exclusivos da resposta tolerante, apenas um (de Função Molecular, GO:0042800: *histone methyltransferase activity* [H3-K4 specific]) esteve entre os dez mais enriquecidos. Da mesma forma, dentre os 81 termos GO enriquecidos e exclusivos da resposta sensível, dois termos de Componente Celular

(GO:0031519: *PcG protein complex*; GO:0031248: *protein acetyltransferase complex*) estiveram dentre os dez mais enriquecidos.

Similarmente, dos termos GO relacionados às *unitags* UR e DR nas respostas TC e SC, 40 foram considerados enriquecidos e compartilhados entre os quatro subconjuntos. Por sua vez, da resposta TC, 25 termos exclusivos foram associados a *unitags* UR, e 14 a *unitags* DR. Já na resposta SC, sete termos exclusivos foram associados a *unitags* UR, e 73 a *unitags* DR (Figura 14b). Os dez termos GO mais enriquecidos de cada grande categoria para alvos associados a *unitags* induzidas e reprimidas nas respostas TC e SC estão descritos nas tabelas 4 e 5, respectivamente.

Figura 14 - Diagramas de Venn da distribuição de termos GO enriquecidos de alvos associados a *unitags* DE nos contrastes tolerante (SD24TvsC [TC]) e sensível (SD24SvsC [SC]) (a) e aqueles associados a UR e/ou DR em ambos os contrastes (b).

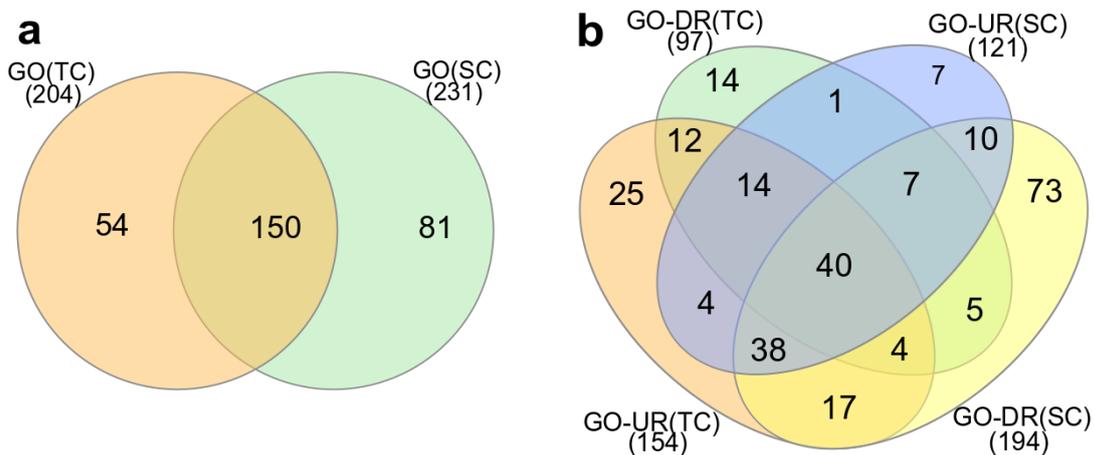


Tabela 4 - Termos GO enriquecidos ( $p < 0,01$ ) associados a *unitags* DE ( $p < 0,05$ ) do contraste tolerante (TC[SD24T vsC]).

Termo id	Termo	Genes alvo	<i>p</i> -value	Categoria
GO:0009755	hormone-mediated signaling pathway	8	9,00E-08	PB
GO:0032870	cellular response to hormone stimulus	8	1.7e-07	PB
GO:0071495	cellular response to endogenous stimulus	8	1.8e-07	PB
GO:0070887	cellular response to chemical stimulus	9	3.2e-07	PB
GO:0016568	chromatin modification	6	5,00E-07	PB
GO:0009734	auxin-activated signaling pathway	5	5.4e-07	PB
GO:0071365	cellular response to auxin stimulus	5	8.3e-07	PB
GO:0071310	cellular response to organic substance	8	9.1e-07	PB
GO:0006325	chromatin organization	6	2.3e-06	PB
GO:0010468	regulation of gene expression	13	4.3e-06	PB
GO:0005634	nucleus	18	2.7e-08	CC
GO:0043229	intracellular organelle	20	0.00061	CC
GO:0043226	organelle	20	0.00063	CC
GO:0005622	intracellular	22	0.00066	CC
GO:0043231	intracellular membrane-bounded organelle	19	0.00116	CC
GO:0043227	membrane-bounded organelle	19	0.00118	CC
GO:0044464	cell part	23	0.00183	CC
GO:0005623	cell	23	0.00206	CC
GO:0044424	intracellular part	21	0.00282	CC
GO:0018024	histone-lysine N-methyltransferase activity	5	1.5e-08	FM
GO:0016278	lysine N-methyltransferase activity	5	4.7e-08	FM
GO:0016279	protein-lysine N-methyltransferase activity	5	4.7e-08	FM
GO:0042054	histone methyltransferase activity	5	4.7e-08	FM
GO:0005515	protein binding	26	7.3e-08	FM
GO:0008276	protein methyltransferase activity	5	2.1e-07	FM
GO:0008170	N-methyltransferase activity	5	5.8e-07	FM
GO:0005488	binding	37	1.4e-06	FM
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	5	6,00E-05	FM
GO:0042800	histone methyltransferase activity (H3-K4 specific)	2	8,00E-05	FM
GO:0009734	auxin-activated signaling pathway	4	6.1e-07	PB
GO:0071365	cellular response to auxin stimulus	4	8.6e-07	PB
GO:0050789	regulation of biological process	11	2,00E-06	PB
GO:0065007	biological regulation	11	6.5e-06	PB
GO:0071310	cellular response to organic substance	5	1.7e-05	PB
GO:0080090	regulation of primary metabolic process	8	3.3e-05	PB
GO:0031323	regulation of cellular metabolic process	8	3.9e-05	PB
GO:0009733	response to auxin	4	4,00E-05	PB
GO:0060255	regulation of macromolecule metabolic process	8	4.2e-05	PB
GO:0070887	cellular response to chemical stimulus	5	4.4e-05	PB
GO:0044798	nuclear transcription factor complex	2	0.00036	CC
GO:0005667	transcription factor complex	2	0.00116	CC
GO:0005634	nucleus	7	0.00197	CC
GO:0055029	nuclear DNA-directed RNA polymerase complex	2	0.00209	CC
GO:0000428	DNA-directed RNA polymerase complex	2	0.0024	CC
GO:0030880	RNA polymerase complex	2	0.0024	CC
GO:1990234	transferase complex	3	0.00372	CC
GO:0061695	transferase complex, transferring phosphorus-containing groups	2	0.004	CC
GO:1902494	catalytic complex	3	0.00934	CC

Genes alvo: quantidade de CoREGs com termos GO enriquecidos associados a *unitags* DEs ( $p < 0,05$ ) UR (vermelho) e DR (azul)

Tabela 5. Termos GO enriquecidos ( $p < 0,01$ ) associados a *unitags* DE ( $p < 0,05$ ) do contraste sensível (SC[SD24SvsC]).

Termo id	Termo	Genes alvo	p-value	Categoria
GO:0006351	transcription, DNA-templated	14	3.6e-08	PB
GO:0097659	nucleic acid-templated transcription	14	4.5e-08	PB
GO:0032774	RNA biosynthetic process	14	4.5e-08	PB
GO:0034654	nucleobase-containing compound biosynthetic process	14	1.8e-07	PB
GO:0009734	auxin-activated signaling pathway	5	2.9e-07	PB
GO:0071365	cellular response to auxin stimulus	5	4.4e-07	PB
GO:0018130	heterocycle biosynthetic process	14	5.7e-07	PB
GO:0009755	hormone-mediated signaling pathway	7	6.7e-07	PB
GO:0019438	aromatic compound biosynthetic process	14	1.1e-06	PB
GO:0032870	cellular response to hormone stimulus	7	1.2e-06	PB
GO:0005634	nucleus	13	1.2e-05	CC
GO:0031248	protein acetyltransferase complex	2	0.00082	CC
GO:1902493	acetyltransferase complex	2	0.00082	CC
GO:0044424	intracellular part	17	0.00317	CC
GO:0005622	intracellular	17	0.00454	CC
GO:0005694	chromosome	3	0.00639	CC
GO:0044464	cell part	18	0.0072	CC
GO:0005623	cell	18	0.00792	CC
GO:0005515	protein binding	23	1.2e-06	FM
GO:0004596	peptide alpha-N-acetyltransferase activity	2	0.00029	FM
GO:0003713	transcription coactivator activity	2	0.00095	FM
GO:0008270	zinc ion binding	9	0.00145	FM
GO:0034212	peptide N-acetyltransferase activity	2	0.00176	FM
GO:0018024	histone-lysine N-methyltransferase activity	2	0.00242	FM
GO:0005488	binding	31	0.00262	FM
GO:0016278	lysine N-methyltransferase activity	2	0.00376	FM
GO:0016279	protein-lysine N-methyltransferase activity	2	0.00376	FM
GO:0042054	histone methyltransferase activity	2	0.00376	FM
GO:0060255	regulation of macromolecule metabolic process	19	1,00E-10	PB
GO:0010468	regulation of gene expression	18	1.1e-10	PB
GO:0019222	regulation of metabolic process	19	1.5e-09	PB
GO:0050789	regulation of biological process	22	3.4e-09	PB
GO:0016568	chromatin modification	7	2.2e-08	PB
GO:0010629	negative regulation of gene expression	8	2.5e-08	PB
GO:0065007	biological regulation	22	3.1e-08	PB
GO:0010605	negative regulation of macromolecule metabolic process	9	5,00E-08	PB
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	7	5.4e-08	PB
GO:0080090	regulation of primary metabolic process	16	7.1e-08	PB
GO:0005634	nucleus	18	2.7e-08	CC
GO:0031519	PcG protein complex	2	1.6e-05	CC
GO:0005622	intracellular	23	4.9e-05	CC
GO:0030015	CCR4-NOT core complex	2	5.5e-05	CC
GO:0070847	core mediator complex	2	7.3e-05	CC
GO:0043229	intracellular organelle	21	9.1e-05	CC
GO:0043226	organelle	21	9.4e-05	CC
GO:0043234	protein complex	10	0.00015	CC
GO:0044424	intracellular part	22	0.00041	CC
GO:0044451	nucleoplasm part	4	0.00048	CC
GO:0003713	transcription coactivator activity	4	1.8e-07	FM
GO:0003712	transcription cofactor activity	5	6.6e-07	FM
GO:0000989	transcription factor activity, transcription factor binding	5	1.2e-06	FM
GO:0000988	transcription factor activity, protein binding	5	2,00E-06	FM
GO:0001104	RNA polymerase II transcription cofactor activity	3	1.4e-05	FM
GO:0001076	transcription factor activity, RNA polymerase II transcription factor binding	3	1.5e-05	FM
GO:0005515	protein binding	23	2,00E-05	FM
GO:0001191	transcriptional repressor activity, RNA polymerase II transcription factor binding	2	4.1e-05	FM
GO:0003677	DNA binding	12	4,00E-04	FM
GO:0003727	single-stranded RNA binding	2	0.0013	FM

Genes alvo: quantidade de CoREGs com termos GO enriquecidos associados a *unitags* DEs ( $p < 0,05$ ) UR (vermelho) e DR (azul)

#### 4.4 ENRIQUECIMENTO DE FTs

Considerando as 74 ESTs com *unitags* DE ( $p < 0,05$ ) da resposta tolerante, foram previstas 666 regulações entre 159 FTs e potenciais promotores de 57 CoREGs (alvos). Dos FTs, dez foram considerados super-representados, envolvendo um total de 90 regulações. Da mesma forma, da resposta sensível foram previstas 1.213 regulações entre 177 FTs e 74 CoREGs alvos. Destes FTs, 28 foram considerados super-representados, envolvendo 481 regulações. Dos FTs super-representados para alvos associados a *unitags* DE da resposta tolerante, a famílias com mais regulações preditas foram C2H2 (dois FTs; 40 regulações) e TALE (um FT; 24 regulações). Similarmente, na resposta sensível, as famílias de FTs com mais regulações preditas foram ERF (13 FTs; 318 regulações), C2H2 (três FTs; 59 regulações) e TALE (um FT; 38 regulações). As famílias de FTs super-representados em números de regulações preditas estão descritos na tabela 6 e representados na figura 15.

Figura 15 - Famílias de FTs super-representados na análise de enriquecimento ( $p$ -value threshold  $\leq 0,01$ ), considerando as *unitags* DE ( $p < 0,05$ ) da resposta tolerante (a) e sensível (b), de cana-de-açúcar (quatro acessos cada), após 24h de supressão de rega.

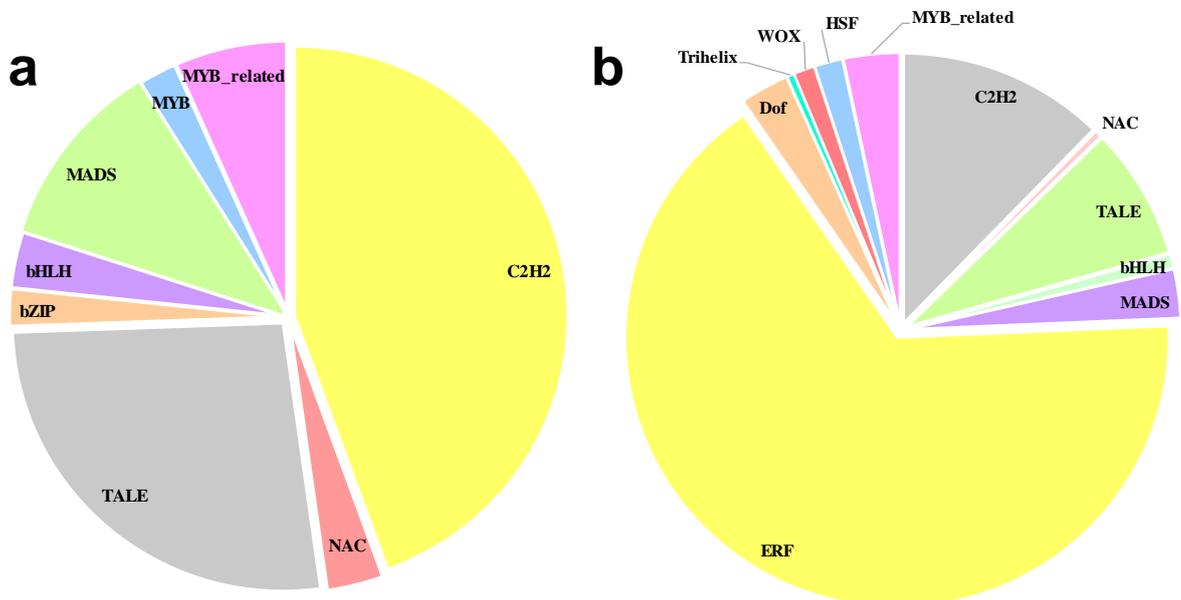


Tabela 6 - Fatores de transcrição super-representados e seu número de regulações preditas para potenciais genes de correguladores associados a *unitags* DE ( $p$ -value < 0,05) da resposta tolerante (SD24TvsC) e sensível (SD24CvsC).

FT	Descrição	Regulações (SD24T vsC)	Regulações (SD24S vsC)
Sobic.008G014801	C2H2	20	27
Sobic.009G024400	C2H2	20	27
Sobic.001G106200	TALE	24	38
Sobic.004G281000	MIKC_MADS	5	6
Sobic.003G034300	MYB_related	6	8
Sobic.002G251700	NAC	3	*
Sobic.003G329500	bZIP	2	*
Sobic.010G051800	bHLH	3	*
Sobic.002G258000	MIKC_MADS	5	*
Sobic.003G352200	MYB	2	*
Sobic.001G486800	ERF	*	28
Sobic.001G034300	Dof	*	14
Sobic.006G167800	ERF	*	29
Sobic.001G297500	MYB_related	*	8
Sobic.002G069166	ERF	*	26
Sobic.009G103800	ERF	*	19
Sobic.009G184300	ERF	*	24
Sobic.003G027000	MIKC_MADS	*	8
Sobic.002G411000	ERF	*	24
Sobic.006G156100	ERF	*	25
Sobic.006G168000	ERF	*	27
Sobic.003G012800	ERF	*	28
Sobic.002G375400	ERF	*	14
Sobic.001G243000	HSF	*	4
Sobic.001G158900	bHLH	*	4
Sobic.006G065800	Trihelix	*	2
Sobic.004G073100	ERF	*	24
Sobic.001G242900	C2H2	*	5
Sobic.002G421800	WOX	*	6
Sobic.008G021800	NAC	*	2
Sobic.001G361100	HSF	*	4
Sobic.004G159500	ERF	*	24
Sobic.009G233100	ERF	*	26

\* fatores de transcrição que não foram preditos para regular alvos em TC ou SC

Na resposta tolerante, 17 famílias de FTs foram associadas a alvos CoREGs com *unitags* UR, sendo as mais representativas ERF (11 FTs), TALE (sete FTs) e GRAS (cinco FTs). Para os alvos CoREGs de *unitags* DR, cinco famílias de FTs se destacaram, sendo as mais representadas TALE (seis FTs) e Dof (quatro FTs). Na resposta sensível, 13 famílias de FTs foram associadas a alvos CoREGs com *unitags* UR, sendo as mais representadas ERF (10 FTs), Dof (10 FTs) e TALE (cinco FTs). Para os alvos CoREGs associados a *unitags* DR, nove famílias de FTs foram identificadas, sendo as mais representadas ERF (14 FTs), TALE (sete FTs) e Dof (cinco FTs).

Adicionalmente, comparando os FTs enriquecidos (mas não super-representados) das duas respostas, observa-se que a maioria dos FTs enriquecidos (147) foram comuns às duas respostas. Estes FTs abrangeram 30 famílias, sendo as mais representadas ERF (29 FTs), NAC (16 FTs) e MYB (10 FTs). A resposta tolerante apresentou 12 FTs enriquecidos exclusivos, de sete famílias: bZIP (três FTs), HD-ZIP (um TF), CPP (um TF), MYB (três FTs), MIKC\_MADS (um TF), SBP (dois FTs) e NAC (um TF). Da mesma maneira, a resposta sensível apresentou 30 FTs exclusivos (Figura 8), de 14 famílias: NAC (cinco FTs), WRKY (quatro FTs), BES1 (dois FTs), ARR-B (dois FTs), bHLH (três FTs), MYB (quatro FTs) SBP (um FT), bZIP (dois FTs), EIL (um FT), WOX (um FT), ARF (um FT), Trihelix (dois FTs), ERF (um FT) e C2H2 (um FT).

#### 4.5 GENÔMICA ESTRUTURAL: PROSPECÇÃO DE *LOCI* ASSOCIADOS A CoREGs EM GENOMAS DE GRAMÍNEAS

Dos alinhamentos BLASTn (*e-value*  $e^{-10}$ ), das ESTs CoREGs com domínios confirmados contra transcritos de sorgo (Sbicolor\_313\_v3.1.transcript), milho (Zmays\_284\_5b+transcript) e arroz (Osativa\_323\_v7.0.transcript), foi possível identificar, por similaridade, cerca de 320 *loci* em sorgo (235 CoREGs similares), 489 em milho (238 CoREGs) e 219 em arroz (196 CoREGs) (Tabela 6). As três espécies, sorgo, milho e arroz, compartilharam 83 ESTs CoREGs (de 19 famílias) associadas a *loci* (Figura 16b). ESTs relacionadas a *loci* exclusivos em cada espécie foram também observados, sendo 71 exclusivas em sorgo (17 famílias CoREG), 69 em milho (14 famílias CoREG) e 54 em arroz (16 famílias CoREG). Outros compartilhamentos de

sorgo incluíram: com milho, 63 ESTs (18 famílias CoREG) e com arroz, 36 (13 famílias CoREG). Milho e arroz compartilharam 23 ESTs associadas a *loci* CoREG de 11 famílias (Figura 16b).

Em sorgo, as famílias CoREG com mais *loci* identificados foram TRAF (64), SET (51) e SNF2 (35). Nos cromossomos, aqueles com maiores números de *loci* observados foram Sb1 (75 *loci*; 56 CoREGs), Sb2 (37 *loci*; 36 CoREGs) e Sb3 (35 *loci*; 32 CoREGs) (Tabela 7; figura 16a). Em milho, as famílias CoREG com mais *loci* foram Aux/IAA (83), SNF2 (54) e *Others* (43), e os cromossomos com mais *loci* foram Zm4 (67 *loci*; 61 CoREGs), Zm6 (67 *loci*; 37 CoREGs), Zm1 (64 *loci*; 39 CoREGs) e Zm2 (60 *loci*; 37 CoREGs) (Tabela 7; figura 16a). Em arroz, as famílias CoREG com mais *loci* foram SNF2 (33), TRAF (33), Aux/IAA (29) e GNAT (19), e os cromossomos com mais *loci* foram Os3 (36 *loci*; 34 CoREGs), Os1 (24 *loci*; 24 CoREGs) e Os6 (23 *loci*; 23 CoREGs) (Tabela 7; figura 16a).

Figura 16 - Quantidade de *loci* associados a correguladores em *S. bicolor*, *Z. mays* e *O. sativa* (a). Diagrama de Venn ilustrando o compartilhamento entre ESTs de cana-de-açúcar associadas a *loci* correguladores entre *S. bicolor*, *Z. mays* e *O. sativa* (b).

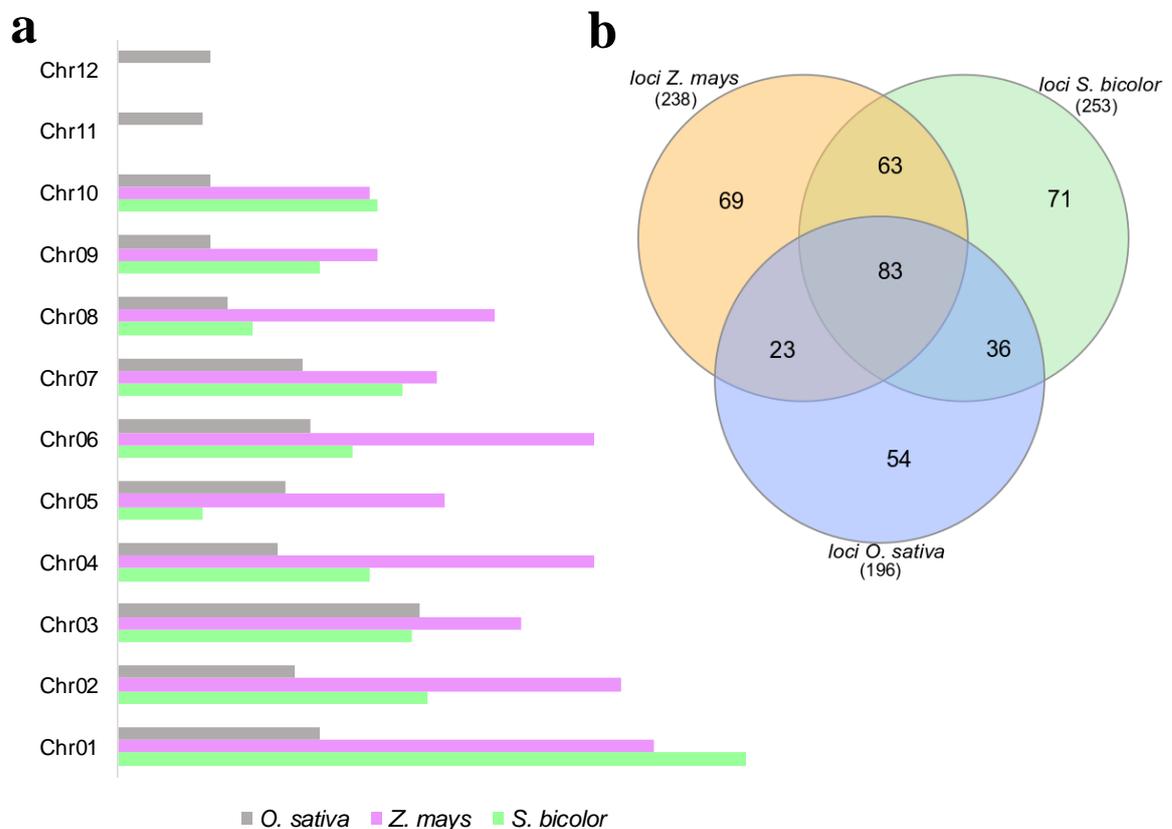


Tabela 7 - Número de *loci* relacionados a CoREGs por espécie (sorgo, milho e arroz) e por família CoREG identificada.

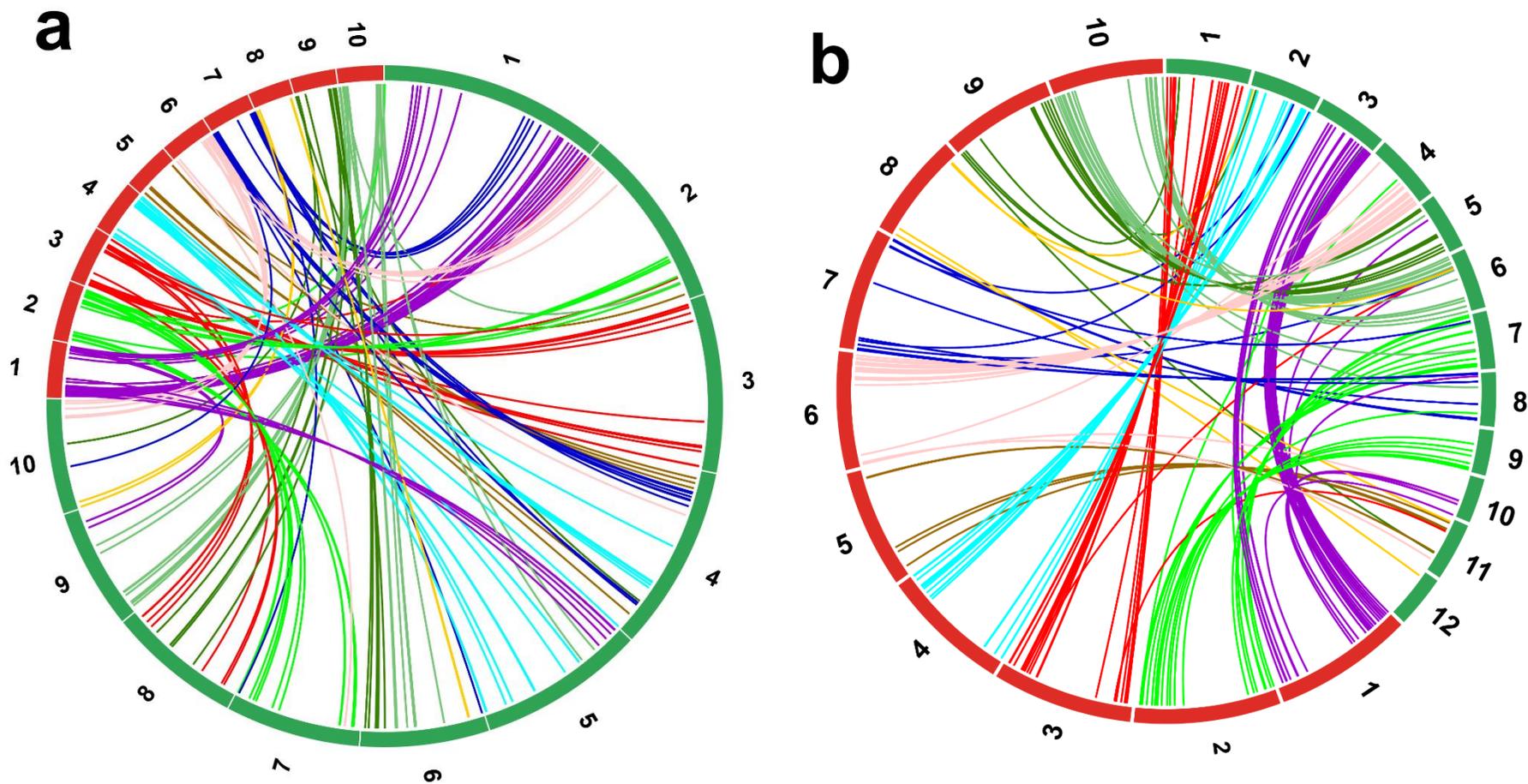
Família CoREG	Número de <i>loci</i>		
	<i>S. bicolor</i>	<i>Z. mays</i>	<i>O. sativa</i>
ARID	5	4	4
AUX/IAA	34	82	29
Coactivator	2	5	3
GNAT	23	34	19
HMG	9	21	8
IWS1	10	15	8
Jumonji	4	8	2
LUG	7	8	6
MBF1	3	16	2
MED6	1	3	1
MED7	1	9	1
mTERF	26	22	14
Others	10	43	10
PHD	14	15	11
Pseudo	2	1	2
RB	4	8	2
Rcd1-like	3	11	3
SET	51	42	19
SNF2	35	54	33
SOH1	2	3	2
SWI/SNF-BAF60b	7	8	6
TAZ	3	3	1
TRAF	64	74	33
<b>Total</b>	<b>320</b>	<b>489</b>	<b>219</b>

#### 4.5.1 Identificação de CoREGs ortólogos em gramíneas

Resultados de *Bidirecional Best-Hit* (BBH) ou recíproco permitiram identificar em sorgo 198 CoREGs ortólogos de cana-de-açúcar. Similarmente, para milho e arroz foram identificados, respectivamente, 187 e 158 ortólogos de genes CoREG de cana-de-açúcar. Análise BLASTp (*e-value*  $e^{-10}$ ) com proteínas dos 198 ortólogos CoREG de sorgo, contra o proteoma de milho (*Zea mays* ID: 12 [assembly B73 RefGen\_v4])

e de arroz (*Oryza sativa* Japonica Group ID: 10 [assembly Build 4.0]) ilustraram os ideogramas circulares da figura 9. Na comparação sorgo com milho, os cromossomos com maiores níveis de microssintenia foram Sb1 e Zm1 (23 *loci*; 13 famílias CoREG), Sb2 e Zm7 (14 *loci*; 12 famílias CoREG), Sb3 e Zm3 (11 *loci*; 10 famílias CoREG) (Figura 17a). Na comparação sorgo com arroz, níveis mais altos de microssintenia envolveram os cromossomos Sb1 e Os3 (23 *loci*; 11 famílias CoREG), Sb3 e Os1 (20 *loci*; 10 famílias CoREG), Sb4 e Os2 (17 *loci*; 11 famílias CoREG) (Figura 17b).

Figura 17 - Ideogramas circulares construídos a partir dos dados de localização cromossômica de genes correguladores ortólogos entre sorgo e cana-de-açúcar nas comparações (via BLASTp  $e\text{-value } e^{-10}$ ) sorgo com milho (a) e sorgo com arroz (b).



#### 4.6 VALIDAÇÃO DA EXPRESSÃO GÊNICA POR RT-qPCR

Resultados de validação pela RT-qPCR, da expressão gênica de cinco genes CoREGs, envolveram diferentes processos: ubiquitinação de proteínas (*TRAF*); transdução de sinal por fosforilação (*phosphorelay*) (*Pseudo-ARR*); degradação de mRNA (*Rcd1-like*); regulação da transcrição dependente de RNA polimerase II (*IWS1*); modificação de histonas (*LUH*). As curvas de dissociação evidenciaram apenas um produto amplificado a partir de cada par de *primer* (Anexo A), enquanto que as curvas geradas, utilizando diluições seriadas do *pool* de cDNAs de raízes das plantas com ou sem estresse, permitiram estimar as eficiências de amplificação (Anexo B), que apresentam variações aceitáveis, dentro do esperado pelo protocolo MIQE (BUSTIN *et al.*, 2009). Os alvos selecionados tiveram suas expressões validadas independentemente para cada cultivar (tolerante e sensível). De forma geral, os resultados de qPCR, validando expressões observadas *in silico* nas análises de HT-SuperSAGE, evidenciaram as diferenças entre as cultivares (Tabela 8). Assim, os resultados para *BPM1* (SD221142;  $FC_{TOL} = 3,179$ ) mostraram indução pelas cultivares tolerantes SP83-2847 e SP83-5073; a indução para *ORR9* (SD6438;  $FC_{TOL} = 2,782$ ) foi validada pela cultivar tolerante SP83-2847; a indução para *Rcd1-like* (SD194342;  $FC_{TOL} = 2,628$ ) foi confirmada pelas cultivares tolerantes SP83-2847 e SP83-5073; a indução para *IWS1* (SD23099;  $FC_{TOL} = 5,563$ ), pelas cultivares tolerantes SP83-2847 e CTC6, e a indução para *LUH* (SD23621;  $FC_{TOL} = 4,027$ ), pela cultivar tolerante SP83-5073.

Tabela 8 - Dados de expressão *in silico* dos candidatos selecionados de *unitags* HT-SuperSAGE de cana-de-açúcar expressas pelos *bulks* de cultivares contrastantes (tolerante e sensível) e respectivas expressões obtidas via RT-qPCR, em raízes após estímulo (24 h de supressão de rega).

<b>Método</b>	<b>TRAF</b>	<b>PseudoARR</b>	<b>Rcd1</b>	<b>LUG</b>	<b>IWS1</b>
<b>HT-SuperSAGE</b>					
<i>Bulk tolerante</i>	3,18* / UR	2,78* / UR	2,62* / UR	4,03* / UR	5,56* / UR
<i>Bulk sensível</i>	1,76* / n.s.	-1,06* / DR	-2,06* / DR	1,05* / n.s.	3,95* / n.s.
<b>RT-qPCR**</b>					
<i>Acessos tolerantes</i>					
CTC6	0,696/ n.s.	0,804/ n.s.	1,786/ n.s.	1,072/ n.s.	2,744/ UR
CTC15	1,206/ n.s.	0,332/ DR	0,787/ n.s.	1,541/ n.s.	0,775/ UR
SP83-2847	1,522/ UR	1,617/ UR	1,685/ UR	1,633/ n.s.	1,914/ UR
SP83-5073	2,802/ UR	1,497/ n.s.	3,784/UR	2,589/ UR	2,839/ UR
<i>Acessos sensíveis</i>					
CTC9	1,337/ n.s.	1,311/ n.s.	0,808/ n.s.	1,230/ n.s.	1,093/ n.s.
CTC13	7,310/ n.s.	7,322/ UR	9,528/UR	8,261/ UR	9,390/ UR
SP90-1638	0,552/n.s.	0,736/ n.s.	0,848/ n.s.	0,755/ n.s.	0,864/n.s.
SP90-3414	1,991/ n.s.	1,417/ n.s.	1,714/ n.s.	1,909/ n.s.	2,557/ UR

## 5 DISCUSSÃO

CoREGs são proteínas que atuam de forma a regular a expressão gênica, porém de maneira indireta, não se ligando as regiões promotoras dos genes-alvo como os fazem os FTs, mas através de interação proteína-proteína (a maior parte das famílias CoREGs atuam desta forma), modificações covalentes nas caudas de histonas, e remodelagem da cromatina. Este trabalho constitui o primeiro ao abordar o CoREGoma (conjunto de CoREGs) de cana-de-açúcar, em raízes após 24 horas da supressão da rega. No geral, não houve diferenças marcantes no total de famílias CoREG associadas às *unitags* DE ( $p\text{-value} < 0,05$ ) detectadas em cada resposta (tolerante: 18 famílias; sensível: 20 famílias). Entretanto, o número de *unitags* DE associadas aos CoREGs foi mais abundante na resposta sensível, ao passo que aquelas de expressão UR foi mais expressiva na resposta tolerante, o que poderia indicar uma maior sensibilidade na detecção e gerenciamento do estresse. Entretanto, os dados justificam uma análise conjunta dos acessos contrastantes, visando obter o máximo de informação das respostas destes acessos frente ao estresse (DA SILVA *et al.*, 2017).

A clusterização hierárquica permitiu identificar um maior número (18) de transcritos CoREGs associados a *unitags* UR ( $p < 0,0001$ ) da resposta tolerante, com modulação pouco expressiva ou mesmo contrastante na resposta sensível. Mesmo padrão foi observado com as *unitags* UR ( $p < 0,0001$ ) da resposta sensível, entretanto, o número dessas *unitags* UR (14) foi menor que aquelas da resposta tolerante (UR), indicando menor mobilização destes na resposta sensível. O inverso foi observado ao se considerar as *unitags* DR ( $p < 0,0001$ ) da resposta tolerante, cujo o número de *unitags* (sete) foi menor que as 14 *unitags* DR ( $p < 0,0001$ ) de modulação pouco expressiva ou contrastante com a resposta tolerante. Se a regulação UR for favorável na resposta ao estresse, os acessos do *bulk* tolerante levariam vantagens no gerenciamento e minimização dos danos decorrentes do estímulo aplicado.

A análise de enriquecimento de termos GO ( $p \leq 0,01$ ), associados aos CoREGs DE ( $p < 0,05$ ) identificou nas duas respostas termos característicos das atividades de um corregulador, como “modificação da cromatina”, “ligação a proteína”, “atividade de histona metiltransferase”, “atividade de co-fator da transcrição”, “núcleo” dentre outros. Outros termos destacados, como exemplo “via de sinalização ativada por auxina”, “via

de sinalização mediada por hormônio”, foram exclusivos da resposta tolerante. Estes termos associados com resposta a fitorregulador, provavelmente, estiveram associados a CoREGs da família Aux/IAA, que deteve o maior número de *unitags* DE da resposta tolerante. Estas proteínas funcionam para reprimir a sinalização da auxina através da dimerização com *ARFs* (*Auxin Responsive Factor*) (TIWARI *et al.*, 2001). A expressão de alguns genes desta classe de proteínas é induzida pela auxina, sendo o seu acúmulo sujeito a *feedback* negativo: Aux/IAA entram em declínio (via complexo SCF<sup>TIR1</sup>) após exposição a auxina, permitindo o acréscimo da transcrição de genes responsivos a auxina e, portanto, garantindo a resposta transiente a este fitoregulador (KORASICK *et al.*, 2014). Os membros desta família foram mais associados a *unitags* DR na resposta tolerante, enquanto que na resposta sensível foram mais UR, exibindo uma divergência de expressão. Outro termo GO enriquecido foi “ligação ao DNA”, que talvez não fosse esperado, considerando que CoREGs não possuem habilidade de se ligar ao DNA. A presença deste termo pode ser explicada pela atuação da família HMG, que foi a segunda mais associada a *unitags* DE nas duas respostas. Estas proteínas atuam modificando a estrutura da cromatina, permitindo que FTs se liguem a promotores e *enhancers*, o que faz com que estas proteínas sejam usualmente consideradas como ‘FTs arquitetônicos’. As HMGs competem com a histona H1 pela ligação ao DNA de ligação para tornar a cromatina acessível a ligação de FTs (REEVES, 2010; MALLIK *et al.*, 2018), daí a presença do termo “ligação ao DNA”.

No geral, os termos GO enriquecidos entre CoREGs com *unitags* da resposta tolerante ou sensível não exibiram muitas diferenças. Entretanto, foi observado um maior número de termos GO exclusivos associados a CoREGs com *unitags* DE da resposta sensível, porém, para este contraste foi observado que a maioria dos termos exclusivos estavam associados a CoREGs com *unitags* DR, enquanto que a resposta tolerante, apesar de apresentar menos termos GO exclusivos, apresentou um maior número de termos associados a CoREGs com *unitags* UR. Logo, a análise de enriquecimento GO confirmou que as anotações para os CoREGs identificados estão condizentes com a atuação do grupo e que a resposta tolerante foi mais associada aos CoREGs induzidos.

A análise de enriquecimento de FTs predisse FTs com mais regulações (e quantidade de CoREGs alvos) associados a *unitags* DE da resposta sensível, bem como, os FTs declarados super-representados. Desses, a família ERF respondeu pela

maior parte (66,11%) das regulações preditas de CoREGs com *unitags* DE, não sendo predita nenhuma regulação de FTs desta família com CoREGs associados a *unitags* DE da resposta tolerante. TFs ERF são exclusivos de plantas e pertencem a grande família AP2/ERF (RIECHMANN *et al.*, 2000). Estas proteínas podem atuar de forma positiva ou negativa, sendo capazes de se ligar a promotores de genes responsivos ao etileno (que incluem GCC box e DRE/CRT), possuindo, ainda, a habilidade de canalizar a sinalização pelo etileno até um conjunto de genes requeridos para processos de desenvolvimento normais, e respostas a estresses abióticos, como seca, salinidade e calor (PIRRELLO *et al.*, 2012, CHENG *et al.*, 2013). Outras famílias de FTs super-representados e exclusivas da resposta sensível foram Dof (DNA binding with one finger), Trihelix (*helix-loop-helix-loop-helix*), HSF (*Heat Stress Transcription Factors*) e WOX (*WUSCHEL-RELATED HOMEODOMAIN*), todas envolvidas em importantes processos de desenvolvimento, como por exemplo controle da morfogênese de órgãos florais, maturação da semente, germinação e também uma série de respostas a estresses abióticos (YANAGISAWA, 2004; GUPTA *et al.*, 2016; DOLZBLASZ *et al.*, 2016; GUO *et al.*, 2016).

Comparando os TFs preditos na regulação de CoREGs associados as *unitags* DE das duas respostas se observou que apesar da resposta sensível ter apresentado maior número de TFs exclusivos, a maior parte (23) destes esteve associada a CoREGs com *unitags* DR, enquanto que dos TFs preditos regulando CoREGs exclusivamente da resposta tolerante, a maioria (09) esteve associada a CoREGs com *unitags* (UR). Isto pode indicar que os corre reguladores são mais requisitados nos acessos tolerantes, ao menos na resposta inicial ao estímulo analisado (24 h de supressão de rega).

A análise de genômica comparativa entre ESTs de CoREGs de cana-de-açúcar com domínios confirmados e transcritos de gramíneas taxonomicamente relacionadas à cana-de-açúcar permitiu identificar genes similares compartilhados entre os três genomas, bem como genes taxonomicamente restritos. Apesar do milho ser mais distante evolutivamente da cana-de-açúcar, quando comparado com sorgo (Soreng *et al.*, 2015), uma maior quantidade (489) de *loci* associados a CoREGs foi observado nesta espécie, provavelmente decorrente do maior tamanho do genoma de milho (média de 2.171.65 Mb), em comparação a sorgo (666.155 Mb) e arroz (382.389 Mb). No entanto, estes *loci* estiveram associados a apenas 238 CoREGs. Alguns

cromossomos apresentaram um maior número de *loci* relacionados, o que também pode ser decorrência do maior tamanho destes cromossomos.

Na identificação de possíveis ortólogos CoREGs, a metodologia BBH (*Bidirecional Best-Hit*) identificou pares de genes que são reciprocamente mais semelhantes entre si ao considerar todos os genes dos organismos (TATUSOV *et al.*, 1997). Assim, sendo ortólogos definidos como genes em espécies com ancestral comum, e que evoluíram apenas através de eventos de especiação, as sequências CoREGs compartilhadas (121) são prováveis ortólogos nestas espécies (TATUSOV *et al.*, 1997; FANG *et al.*, 2010). Conforme esperado, sorgo apresentou mais genes ortólogos com cana-de-açúcar (198), quando comparado com milho (187) ou arroz (158), que são mais distantes taxonomicamente. Além dos potenciais ortólogos, sequências CoREGs restritas as espécies analisadas (sorgo [18]; milho [33]; arroz [15]) podem indicar eventos de duplicações gênicas ao longo do tempo evolutivo. Por sua vez, a maior microssintenia observada entre sorgo e arroz, pode ser devida as quantidades similares de eucromatina presente nos genomas destas espécies, bem como a densidade gênica e a ordem dos genes (PATERSON *et al.*, 2009). Ainda, a abundância de *loci* observada em alguns cromossomos, bem como o agrupamento de *loci* apresentado nas extremidades dos cromossomos de sorgo e alguns de milho (Zm2, Zm3 e Zm7) e arroz (Os6, Os4 e Os6), pode indicar potenciais regiões de QTLs (*Quantitative Trait Loci*), úteis para exploração em mapeamentos genéticos e seleção assistida por marcadores (KIDO *et al.*, 2013).

## 5.1 PANORAMA DA EXPRESSÃO POR qPCR DE ALVOS CoREGs E PROVÁVEIS ATUAÇÕES EM PLANTAS SOB ESTRESSE

- **Degradação de proteínas: via ubiquitinação (TRAF-BPM1)**

A ubiquitinação de proteínas é uma das modificações pós-traducionais mais relevantes na regulação de uma variedade de processos intracelulares, como interação proteína-proteína, regulação alostérica e degradação proteossomal. Um gene que codifica um adaptador substrato específico do complexo E3 ubiquitin-protein ligase (CUL3-RBX1-BTB), gene *BPM1* (MATH domain-containing protein 1), que apresenta domínios BTB/POZ, e que medeia a degradação de proteínas alvo, teve a expressão validada por RT-qPCR, e a indução confirmada para as cultivares

tolerantes SP83-2847 e SP83-5073. O domínio BTB/POZ é altamente conservado, e presente em eucariotos e alguns vírus (AHMAD *et al.*, 1998). Outros domínios característicos de interação proteína-proteína, tais como ARMADILLO, ANKYRIN e MATH (meprin e TRAF [receptor associado ao fator de necrose tumoral] homólogo) podem ser encontrados combinados com BTB/POZ, evidenciando atuações variadas em processos biológicos (WEBER *et al.*, 2005). Em plantas de *Arabidopsis* essas proteínas interagiram, via ensaio duplo híbrido, com FTs de resposta ao etileno (ERF/AP2), um regulador de respostas a estresses ambientais (WEBER e HALLMANN, 2009). Os autores levantaram que as BPMs (proteínas BPM1 e BPM4, as quais foram exclusivamente induzidas pela seca, em culturas em placas de Petri expostas ao ar por 6 h) se ligaram a ERF/AP2, influenciando na ligação deste ao DNA, e posteriormente acarretando na degradação via proteassomo. Também em plantas de *Arabidopsis*, LECHNER *et al.*, (2011) observaram a interação, via ensaio duplo híbrido, de BPMs com o FT ATHB6 (zíper de leucina homeobox classe I), um regulador negativo de resposta ao ABA e chave nas respostas de plantas a estresses abióticos. Plantas com BPMs silenciadas (via inserções T-DNA e miRNAs artificiais) e de expressões reduzidas mostraram alterações fenotípicas de crescimento e reprodução. Quando a função do complexo CUL3<sup>BPM</sup> foi diminuída, através de um inibidor de proteassomo, houve acúmulo de ATHB6 e redução do crescimento e fertilidade, bem como, comprometimento do comportamento dos estômatos, evidenciando atuação em células guarda, durante condições de estresse.

- **Transdução de sinal por fosforilação (Pseudo ARR-B/RR9)**

Proteínas ORR9 (RR9) fazem parte do chamado sistema de dois componentes, que é constituído por uma histidina quinase (HK – que recebe o sinal) e um regulador de resposta (RR – que amplia o sinal via fosforilação His->Asp, e regula a expressão dos genes alvo) (HWANG *et al.*, 2002). No presente trabalho, um potencial gene regulador de resposta *ORR9* (two-component response regulator ORR9), teve a expressão validada por RT-qPCR, com indução pela cultivar tolerante SP83-2847. Genes do sistema de dois componentes há tempos têm sido associados a respostas a estresses abióticos, atuando em redes de regulação gênica associadas a FTs, como MYB (myeloblastosis) e ERF (*ethylene-responsive element binding factor*) (RASHOTTE *et al.*, 2006; DORTAY *et al.*, 2007). Relacionados ao sistema de dois componentes, THU *et al.*, (2015) avaliaram por qPCR a expressão de genes *HK* (11)

e *RR* (15), em raízes de dois genótipos de soja contrastantes a seca [tolerante (DT51) e sensível (MTD720)], sob supressão de rega por 15 dias, e plantas regadas diariamente. Quatro reguladores de resposta (*GmRR04*, 17, 34, e *GmPRR39*) exibiram altos níveis de expressão na cultivar tolerante em relação a sensível. Os autores destacaram uma provável atuação destes genes na tolerância de DT51, e uma correlação positiva entre o número de genes associados a este sistema e induzidos pela seca e a capacidade de tolerância. A busca por elementos *cis* nas regiões promotoras dos genes estudados detectou elementos responsivos a seca e ao ABA, que incluíram ABRE1 (*ABA responsive element 1*), ABRE2 (*ABA responsive element 2*), MYBR (*MYB recognition site*), MYCR (*MYC recognition site*) e NACR (*NAC recognition site*). Em plantas de *Cicer arietinum*, AHMAD *et al.*, (2019) observaram expressão aumentada (qPCR) de dois reguladores de resposta (*CarRR2* e *CarRR12*), em folhas de plantas submetidas a supressão de rega por 10 dias, em relação as plantas irrigadas, evidenciando o envolvimento destes em resposta a seca.

- **Degradação de mRNA (Rcd1-like: CCR4-NOT transcription complex subunit 9)**

Proteínas CCR4 compõem o complexo CCR4-NOT, que atua no processo de deadenilação de mRNAs (encurtamento da cauda poli-A), conduzindo a molécula a subsequente degradação, via atividade de enzimas em complexos de *deccaping* e exosomo, constituindo assim em uma importante etapa na regulação da expressão gênica em eucariotos (THARUN e PARKER 2001; BELOSTOTSKY e SIEBURTH 2009). Duas unidades catalíticas constituem o complexo CCR4-NOT: CCR4 (*carbon catabolite repressor 4*) e CAF1 (*CCR4 associated factor 1*). Em arroz, CAF1 interage com NOT1, conectando-o com CCR4, garantindo a função do complexo (CHOU *et al.*, 2017). No presente trabalho, um provável gene codificador de subunidade do complexo CCR4-NOT mostrou-se induzido (RT-qPCR) pelas cultivares tolerantes SP83-2847 e SP83-5073 após o tratamento aplicado. Componentes do complexo CCR4-NOT já foram associados com respostas a estresses em plantas. Em *Citrus sinensis*, SHIMO *et al.*, (2019) verificaram via qPCR que a expressão do gene *CsCAF1* foi induzida após 24 e 48 h de inoculação de *Xanthomonas aurantifolii* em folhas. O aumento de expressão foi correlacionado com o decréscimo na expressão de *CsLOB1* (*Lateral Organ Boundaries 1*), o gene para susceptibilidade a necrose em citrus. Os autores observaram também que um inibidor da atividade de *CsCAF1* (derivado de

pirazolo [4,3-d]-pirimidina) ocasionou um aumento significativo na formação de úlceras nas folhas da planta. Em *Arabidopsis*, a proteína *AtCCR4* interagiu, via BiFC, com proteínas Pumilio (*APUM5* e *APUM2*), as quais possuem domínio de ligação ao RNA, não existente em *AtCCR4* (ARAE *et al.*, 2019). *APUM5* foi reportada em respostas a estresses abióticos. HUH e PAEK (2014) relataram que o promotor de *APUM5* possui elementos de resposta ao ABA (*ABFs binding motif*, *ABREATRD22*) e elementos de reconhecimento de MYB (*MYB4 binding motif*, *MYB1AT*), comumente encontrados em genes estresse-responsivos. Os mesmos autores também observaram que plantas de *Arabidopsis* superexpressando *APUM5* (*35S-APUM5*) exibiam fenótipo hipersensível a seca (irrigação suspensa por 14 dias), ao sal (150 mM de NaCl) e ABA exógeno (0.5  $\mu$ M e 0.7  $\mu$ M), enquanto que as plantas nocauteadas (*APUM5-RNAi*) foram mais tolerantes a estes tratamentos quando comparadas as plantas selvagens (*Col-0*).

- **Regulação da transcrição dependente de RNA polimerase II (*IWS1*)**

A família *IWS1* compreende o chamado complexo mediador, que em plantas possui 34 subunidades, e que atua como um coativador na regulação de quase todos os genes dependentes da maquinaria transcricional de RNA pol II. As proteínas relacionadas formam uma ponte entre a maquinaria da transcrição e outras proteínas regulatórias específicas, auxiliando na montagem do complexo de pré-iniciação (CONAWAY e CONAWAY, 2011). No presente trabalho, um provável gene para a subunidade 26c do complexo mediador teve a indução validada pelas cultivares tolerantes SP83-2847 e CTC6 após a supressão de rega. Alguns componentes do complexo mediador já foram associados a estresses ambientais. HEMSLEY *et al.*, (2014) avaliaram o papel de *SFR6/MED16* na expressão de genes responsivos ao frio. Os autores observaram em plantas de *Arabidopsis* superexpressando *CBF1* (*C-repeat binding factor 1*), que induz a expressão de genes regulados pelo frio (*COR*), que os genes *KIN2* e *GOLS3* (alvos de *CBF1*) tinham expressão constitutiva em plantas selvagens *Col-0*, e não foram expressos no mutante nulo *sfr6-1*. Os autores também mostraram que os mutantes *sfr6-1*, *med2* e *med14* falharam na aclimação ao frio (duas semanas a 5 °C) quando comparadas as plantas selvagens, indicando que estas plantas provavelmente não expressavam genes *COR*, e que estas subunidades do complexo mediador seriam requeridas durante a aclimação ao frio. *PTF1/MED25* também é uma subunidade do complexo mediador e está associada ao desenvolvimento de raízes, fator determinante na aquisição de água e

consequentemente de tolerância a estresses. RAYA-GONZALEZ *et al.*, (2014) verificaram que plantas de *Arabidopsis* superexpressando MED25 tinham raízes primárias menores que a planta selvagem, enquanto que os mutantes *ptf1* exibiam aumento no número e tamanho de raízes laterais. Eles observaram também que a expressão do marcador responsivo a auxina (*DR5:uidA*) foi mais forte em *ptf1*, indicando que MED25 controla a formação de raízes por regular a sinalização da auxina, fitorregulador que atua diretamente neste processo. SUNDARAVELPANDIAN *et al.*, (2013) também mostraram que mutantes de *Arabidopsis ptf1* tiveram as peroxidases da classe III altamente reprimidas (dados de *microarray* confirmados por qPCR), e mostraram também um decréscimo no nível de  $H_2O_2$  e um acréscimo de  $O_2^-$  nas raízes de *ptf1* em relação a planta selvagem, evidenciando que PTF1/MED25 atua mantendo o balanço de ROS em raízes de *Arabidopsis*.

- **Modificação de histonas: (*transcriptional corepressor LEUNIG\_HOMOLOG-like* [LUH])**

As proteínas LUH (homólogas de LUG) são membros da família de COREGs transcricionais Gro/Tup1, as quais não possuem capacidade de ligação ao DNA, mas exercem suas funções recrutando proteínas específicas via interação proteína-proteína (COUREY e JIA 2001; LEE e GOLZ, 2012). LUH interage, via domínio LUFS, com as adaptadoras SEUSS-like1 (SKL1) e SKL2 para inibir a transcrição (SHRESTHA *et al.*, 2014). No presente trabalho, um provável gene *LUH* teve sua expressão validada por qPCR, sendo induzido pela cultivar tolerante SP83-5073 após a supressão de rega. A atuação de proteínas LUH frente a estresses ambientais tem sido reportada. SHRESTHA *et al.*, (2014) verificaram que plantas mutantes de *Arabidopsis (luh, slk1 e slk2)* exibiam fenótipo de tolerância quando submetidas ao sal (125 mM de NaCl) ou estresse osmótico (300 mM de manitol), em comparação com as plantas selvagens. Também confirmaram que LUH medeia a repressão através do recrutamento de HDACs aos genes alvo, de forma que em um ensaio com protoplastos na presença de TSA (inibidor da atividade de HDAC) LUH falhou em reprimir a expressão do gene repórter. Os mutantes também mostraram elevados níveis de expressão (por qPCR) dos genes *RD20*, *MYB2* e *NAC019*, importantes em tolerância a estresses abióticos. Estes genes alvo exibiram elevados níveis de acetilação na histona H3 (Lys-9 e Lys-14) nos mutantes *luh*, *slk1* e *slk2*, quando comparados as plantas selvagens, reforçando o mecanismo de repressão por

recrutamento de HDACs. GENG *et al.*, (2017) observaram que plantas *Arabidopsis* mutantes *luh* tinham o crescimento de suas raízes pouco inibido na presença de concentrações graduais de Al e exibiram baixa acúmulo de Al nas suas raízes, indicando que LUH é um regulador negativo de resistência ao Al. Os autores também avaliaram o perfil transcricional dos mutantes *luh* e das plantas selvagens utilizando a técnica RNA-Seq, tendo identificados genes diferencialmente expressos em raízes de *luh* tratadas com Al em comparação com as plantas selvagens tratadas. Em análise de ontologia gênica, os autores destacaram genes associados a modificação da parede celular. Ao mineraram os dados para genes associados a PME (*pectin methylesterase*), determinantes da sensibilidade ao Al, observaram que a expressão por qPCR de candidato *PME46* foi maior nos mutantes *luh* em comparação a planta selvagem.

Logo, os dados aqui analisados provavelmente refletem uma melhor performance dos acessos que compõem o *bulk* tolerante em relação ao *bulk* sensível, em decorrência do estímulo aplicado (24 h de supressão de rega). A validação da expressão dos genes propostos (escolhidos levando-se em consideração genes DEs  $p < 0,0001$ ), mostrou que nem todas as cultivares do mesmo *bulk* apresentaram as mesmas respostas. Os diferentes fundos genéticos das cultivares e particularmente a natureza complexa da tolerância a seca contribuem fortemente para estas diferenças (HU e XIONG, 2014). Os candidatos propostos são promissores ao desenvolvimento de marcadores moleculares funcionais e auxiliando os melhoristas a identificar acessos com melhores respostas frente a seca.

## 6 CONCLUSÕES

O presente trabalho representa o primeiro CoREGoma de cana-de-açúcar a abordar a expressão de transcritos relacionados e expressos por cultivares tolerante

e sensíveis ao déficit hídrico, todas oriundas de programa de melhoramento de cana-de-açúcar de reconhecimento mundial (Centro de Tecnologia Canavieira, SP, Brasil), após um estímulo de 24 h de supressão de rega. O CoREGoma abrangiu representantes de 24 famílias, sendo a família AUX/IAA a mais representada. Essa é também uma das mais importantes e reportadas famílias em respostas de plantas a estresses ambientais. A associação das ESTs CoREGs de cana-de-açúcar com as *unitags* HT-SuperSAGE expressas após o estímulo mostrou mais de uma centena de *unitags* diferencialmente expressas em cada resposta, mas aquelas superexpressas de forma significativa (UR) foram predominantes na resposta tolerante. A atividade de correulador da expressão gênica dos CoREGs foi ressaltada nos termos GO enriquecidos. Dos FTs preditos a regular CoREGs exclusivamente da resposta tolerante, a maioria foi associada a *unitags* UR do *bulk* tolerante, ao passo que dos FTs preditos a regular CoREGs exclusivamente da resposta sensível, a maioria esteve associada a *unitags* DR desta resposta. No mapeamento *in silico* de *loci* CoREGs, estes foram observados distribuídos ao longo de todos os cromossomos das gramíneas sorgo, milho e arroz, espécies evolutivamente relacionadas com a cana-de-açúcar (família Poaceae). Em vista da maior sintenia e proximidade taxonômica, um maior número de CoREGs entre cana-de-açúcar e sorgo foram observados. No entanto, considerando os prováveis ortólogos de sorgo, tendo por base os candidatos expressos de cana-de-açúcar, e as comparações de sorgo com milho e sorgo com arroz, mesmo sendo arroz mais distante de sorgo, em comparação a sorgo com milho, foram observados maior grau de microssintênias específicas na comparação sorgo com arroz. Ainda, em todos os cromossomos de sorgo e alguns de milho e arroz, os CoREGs estiveram quase que totalmente associados as extremidades dos braços cromossômicos. Este painel de genes e transcritos CoREGs expressos podem ser úteis em estratégias biotecnológicas junto ao melhoramento genético da cultura frente a seca, na seleção de acessos superiores frente ao referido estresse. Estes genes também podem ser manipulados para avaliar seus efeitos em melhorar a tolerância a seca seja por superexpressão ou nocaute.

## REFERÊNCIAS

- ARAE, Toshihiro et al. Identification of Arabidopsis CCR4-NOT complexes with Pumilio RNA-binding proteins, APUM5 and APUM2. **Plant and Cell Physiology**, v. 60, n. 9, p. 2015-2025, 2019.
- ADAMS, Mark D. et al. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**, v. 252, n. 5013, p. 1651-1656, 1991.
- AUDIC, Stéphane; CLAVERIE, Jean-Michel. The significance of digital gene expression profiles. **Genome research**, v. 7, n. 10, p. 986-995, 1997.
- AHMAD, Bilal et al. Genome-wide identification and expression analysis of two component system genes in *Cicer arietinum*. **Genomics**, v. 112, n. 2, p. 1371-1383, 2020.
- AHMAD, K. Farid; ENGEL, Christian K.; PRIVÉ, Gilbert G. Crystal structure of the BTB domain from PLZF. **Proceedings of the National Academy of Sciences**, v. 95, n. 21, p. 12123-12128, 1998.
- ASHTON, A. R. et al. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997). **Nucl Acids Res**, v. 25, p. 3389-3402, 1997.
- ALTSCHUL, Stephen F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403-410, 1990.
- AMALRAJ, V. A.; BALASUNDARAM, N. On the taxonomy of the members of ' . 2006.
- BAILEY, TL, BodenM et al. MEME suite: tools for motif Discovery and searching. **nucleic AcidsResearch**, v. 37, p. W202rW208, 2009.
- BARTLETT, Karin A. Serial analysis of gene expression. In: **PCR Protocols**. Humana Press, 2003. p. 271-283.
- BELOSTOTSKY, Dmitry A.; SIEBURTH, Leslie E. Kill the messenger: mRNA decay and plant development. **Current opinion in plant biology**, v. 12, n. 1, p. 96-102, 2009.
- BERGER, Shelley L. The complex language of chromatin regulation during transcription. **Nature**, v. 447, n. 7143, p. 407-412, 2007.
- BEZERRA, Tais Lacerda; RAGAUSKAS, Art J. A review of sugarcane bagasse for second-generation bioethanol and biopower production. **Biofuels, Bioproducts and Biorefining**, v. 10, n. 5, p. 634-647, 2016.
- BODNER, G.; ALIREZA, A.; KAUL, H. P. Management of crop water under drought: A review. *Agron. Sustain. Dev*, v. 35, 2015.
- BONNETT, Graham D. Developmental stages (phenology). **Sugarcane: physiology, biochemistry, and functional biology**, p. 35-53, 2013.
- BOYKO, Alex; KOVALCHUK, Igor. Epigenetic control of plant stress response. **Environmental and molecular mutagenesis**, v. 49, n. 1, p. 61-72, 2008.
- BLUM, Abraham. Drought resistance, water-use efficiency, and yield potential—are they compatible, dissonant, or mutually exclusive?. **Australian Journal of Agricultural Research**, v. 56, n. 11, p. 1159-1168, 2005.

- BURDO, Brett et al. The Maize TFome—development of a transcription factor open reading frame collection for functional genomics. **The Plant Journal**, v. 80, n. 2, p. 356-366, 2014.
- BUSTIN, Stephen A. et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. 2009.
- BUSZEWICZ, Daniel et al. HD2C histone deacetylase and a SWI/SNF chromatin remodelling complex interact and both are involved in mediating the heat stress response in Arabidopsis. **Plant, cell & environment**, v. 39, n. 10, p. 2108-2122, 2016.
- CHEN, Li-Ting et al. Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response. **Journal of experimental botany**, v. 61, n. 12, p. 3345-3353, 2010. (a)
- CHEN, Li-Ting; WU, Keqiang. Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. **Plant signaling & behavior**, v. 5, n. 10, p. 1318-1320, 2010. (b)
- CHEN, Z. Jeffrey; TIAN, Lu. Roles of dynamic and reversible histone acetylation in plant development and polyploidy. **Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression**, v. 1769, n. 5-6, p. 295-307, 2007.
- CHENG, Mei-Chun et al. The Arabidopsis ETHYLENE RESPONSE FACTOR1 regulates abiotic stress-responsive gene expression by binding to different cis-acting elements in response to different stress signals. **Plant physiology**, v. 162, n. 3, p. 1566-1582, 2013.
- CHEVALIER, David; MORRIS, Erin R.; WALKER, John C. 14-3-3 and FHA domains mediate phosphoprotein interactions. **Annual review of plant biology**, v. 60, p. 67-91, 2009.
- Companhia Nacional de Abastecimento (CONAB). Acompanhamento de safra brasileira: cana-de-açúcar, levantamento dezembro/2019. Disponível em: <<https://www.conab.gov.br/info-agro/safras/cana>> acesso em: 08 Jan, 2020.
- CLAPIER, Cedric R.; CAIRNS, Bradley R. The biology of chromatin remodeling complexes. **Annual review of biochemistry**, v. 78, p. 273-304, 2009.
- AUDIC, Stéphane; CLAVERIE, Jean-Michel. The significance of digital gene expression profiles. **Genome research**, v. 7, n. 10, p. 986-995, 1997.
- CONAWAY, Ronald C.; CONAWAY, Joan Weliky. Function and regulation of the Mediator complex. **Current opinion in genetics & development**, v. 21, n. 2, p. 225-230, 2011.
- CONESA, Ana; GÖTZ, Stefan. Blast2GO: a comprehensive suite for functional analysis in plant genomics. **International journal of plant genomics**, v. 2008, 2008.
- CORONA, Davide FV; TAMKUN, John W. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. **Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression**, v. 1677, n. 1-3, p. 113-119, 2004.
- CORPET, Florence. Multiple sequence alignment with hierarchical clustering. **Nucleic acids research**, v. 16, n. 22, p. 10881-10890, 1988.

- COSMA, Maria Pia; TANAKA, Tomoyuki; NASMYTH, Kim. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle–and developmentally regulated promoter. **Cell**, v. 97, n. 3, p. 299-311, 1999.
- COUREY, Albert J.; JIA, Songtao. Transcriptional repression: the long and the short of it. **Genes & development**, v. 15, n. 21, p. 2786-2796, 2001.
- CHOU, Wei-Lun et al. Novel interaction between CCR4 and CAF1 in rice CCR4–NOT deadenylase complex. **Plant molecular biology**, v. 93, n. 1-2, p. 79-96, 2017.
- CRAMER, Grant R. Abiotic stress and plant responses from the whole vine to the genes. **Australian Journal of Grape and Wine Research**, v. 16, p. 86-93, 2010.
- DA SILVA, Manassés Daniel et al. Genotype-dependent regulation of drought-responsive genes in tolerant and sensitive sugarcane cultivars. **Gene**, v. 633, p. 17-27, 2017.
- DAI, Xinbin; ZHUANG, Zhaohong; ZHAO, Patrick Xuechun. psRNATarget: a plant small RNA target analysis server (2017 release). **Nucleic acids research**, v. 46, n. W1, p. W49-W54, 2018.
- DANIELS, J.; ROACH, B. T.; HEINZ, D. J. Sugarcane improvement through breeding. **Elsevier, Amsterdam, 1987**. [Google Scholar], 1987.
- DE MORAIS, Lizz Kezzy et al. Breeding of sugarcane. In: **Industrial Crops**. Springer, New York, NY, 2015. p. 29-42.
- DING, Yong; AVRAMOVA, Zoya; FROMM, Michael. The Arabidopsis trithorax-like factor ATX1 functions in dehydration stress responses via ABA-dependent and ABA-independent pathways. **The Plant Journal**, v. 66, n. 5, p. 735-744, 2011.
- DOLZBLASZ, Alicja et al. Stem cell regulation by Arabidopsis WOX genes. **Molecular plant**, v. 9, n. 7, p. 1028-1039, 2016.
- DONG, ZhiCheng; CHEN, Yan. Transcriptomics: advances and approaches. **Science China Life Sciences**, v. 56, n. 10, p. 960-967, 2013.
- DORTAY, Hakan et al. Toward an interaction map of the two-component signaling pathway of Arabidopsis thaliana. **Journal of proteome research**, v. 7, n. 9, p. 3649-3660, 2008.
- EISEN, Michael B. et al. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**, v. 95, n. 25, p. 14863-14868, 1998.
- Food and Agriculture Organization of the United Nations (FAO). Disponível em: <[www.fao.org/faostat](http://www.fao.org/faostat)> acesso em: 18 Dez, 2018.
- FANG, Gang et al. Getting started in gene orthology and functional analysis. **PLoS computational biology**, v. 6, n. 3, 2010.
- FANG, Hui et al. Expression analysis of histone acetyltransferases in rice under drought stress. **Biochemical and biophysical research communications**, v. 443, n. 2, p. 400-405, 2014.

- FENG, Da-Fei; DOOLITTLE, Russell F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. **Journal of molecular evolution**, v. 25, n. 4, p. 351-360, 1987.
- FERDOUS, Jannatul; HUSSAIN, Syed S.; SHI, Bu-Jun. Role of micro RNA s in plant drought tolerance. **Plant biotechnology journal**, v. 13, n. 3, p. 293-305, 2015.
- FU, Wenqun; WU, Keqiang; DUAN, Jun. Sequence and expression analysis of histone deacetylases in rice. **Biochemical and biophysical research communications**, v. 356, n. 4, p. 843-850, 2007.
- FULLER, Jonathan C. et al. Biggest challenges in bioinformatics. **EMBO reports**, v. 14, n. 4, p. 302-304, 2013.
- GAO, Jianjiong; XU, Dong. The Musite open-source framework for phosphorylation-site prediction. In: **BMC bioinformatics**. BioMed Central, 2010. p. S9.
- GENG, Xiaoyu et al. LEUNIG \_ HOMOLOG transcriptional co-repressor mediates aluminium sensitivity through PECTIN METHYLESTERASE 46-modulated root cell wall pectin methylesterification in Arabidopsis. **The Plant Journal**, v. 90, n. 3, p. 491-504, 2017.
- GENTILE, Agustina et al. MicroRNAs and drought responses in sugarcane. **Frontiers in plant science**, v. 6, p. 58, 2015.
- GHANASHYAM, Krishna et al. Effects of thermal annealing on the properties of zirconia films prepared by ion-assisted deposition. **Materials Science and Engineering: B**, v. 21, n. 1, p. 10-13, 1993.
- GODA, Hideki et al. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. **The Plant Journal**, v. 55, n. 3, p. 526-542, 2008.
- GRABHERR, M. G. Haas bJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, raychowdhury r, Zeng Q, et al. 2011. Full-length transcriptome assembly from rNA-seq data without a reference genome. **Nat biotech**, v. 29, n. 7, p. 644-652.
- GRIVET, Laurent et al. A review of recent molecular genetics evidence for sugarcane evolution and domestication. **Ethnobotany Research and Applications**, v. 2, p. 009-017, 2004.
- GUO, Liquan et al. Cullin-RING Ubiquitin Ligase Family in Plant Abiotic Stress Pathways F. **Journal of integrative plant biology**, v. 55, n. 1, p. 21-30, 2013.
- GUO, Meng et al. The plant heat stress transcription factors (HSFs): structure, regulation, and function in response to abiotic stresses. **Frontiers in plant science**, v. 7, p. 114, 2016.
- GUPTA, Shubhra et al. Molecular cloning and expression profiling of multiple Dof genes of Sorghum bicolor (L) Moench. **Molecular biology reports**, v. 43, n. 8, p. 767-774, 2016.
- HAN, Soon-Ki et al. Roles and activities of chromatin remodeling ATP ases in plants. **The Plant Journal**, v. 83, n. 1, p. 62-77, 2015.

- HAN, Soon-Ki; WAGNER, Doris. Role of chromatin in water stress responses in plants. **Journal of experimental botany**, v. 65, n. 10, p. 2785-2799, 2014.
- HARK, Amy T. et al. Two Arabidopsis orthologs of the transcriptional coactivator ADA2 have distinct biological functions. **Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms**, v. 1789, n. 2, p. 117-124, 2009.
- HEBERLE, Henry et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. **BMC bioinformatics**, v. 16, n. 1, p. 169, 2015.
- HEMSLEY, Piers A. et al. The Arabidopsis mediator complex subunits MED16, MED14, and MED2 regulate mediator and RNA polymerase II recruitment to CBF-responsive cold-regulated genes. **The Plant Cell**, v. 26, n. 1, p. 465-484, 2014.
- HIETER, Philip; BOGUSKI, Mark. Functional genomics: it's all how you read it. **Science**, v. 278, n. 5338, p. 601-602, 1997.
- HIGGINS, Desmond G.; SHARP, Paul M. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. **Gene**, v. 73, n. 1, p. 237-244, 1988.
- HOGEWEG, Paulien; HESPER, Ben. The alignment of sets of sequences and the construction of phyletic trees: an integrated method. **Journal of molecular evolution**, v. 20, n. 2, p. 175-186, 1984.
- HOLLENDER, Courtney; LIU, Zhongchi. Histone deacetylase genes in Arabidopsis development. **Journal of integrative plant biology**, v. 50, n. 7, p. 875-885, 2008.
- HU, Honghong; XIONG, Lizhong. Genetic engineering and breeding of drought-resistant crops. **Annual review of plant biology**, v. 65, p. 715-741, 2014.
- HU, Zhaorong et al. Histone acetyltransferase GCN 5 is essential for heat stress-responsive gene activation and thermotolerance in Arabidopsis. **The Plant Journal**, v. 84, n. 6, p. 1178-1191, 2015.
- HUANG, Xiaoqi. On global sequence alignment. **Bioinformatics**, v. 10, n. 3, p. 227-235, 1994.
- HUH, Sung Un; PAEK, Kyung-Hee. APUM5, encoding a Pumilio RNA binding protein, negatively regulates abiotic stress responsive gene expression. **BMC plant biology**, v. 14, n. 1, p. 75, 2014.
- HWANG, Ildoo; CHEN, Huei-Chi; SHEEN, Jen. Two-component signal transduction pathways in Arabidopsis. **Plant Physiology**, v. 129, n. 2, p. 500-515, 2002.
- INMAN-BAMBER, N. G.; SMITH, D. M. Water relations in sugarcane and response to water deficits. **Field crops research**, v. 92, n. 2-3, p. 185-202, 2005.
- KANEHISA, Minoru et al. The KEGG database. In: **Novartis Foundation Symposium**. Chichester; New York; John Wiley; 1999, 2002. p. 91-100.
- KAHL, Günter et al. Reduced representation sequencing of plant stress transcriptomes. **Journal of Plant Biochemistry and Biotechnology**, v. 21, n. 1, p. 119-127, 2012.
- KIDO, Ederson A. et al. Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. **BMC bioinformatics**, v. 14, n. 1, p. S7, 2013.

- KIDO, Ederson Akio et al. New insights in the sugarcane transcriptome responding to drought stress as revealed by supersage. **The Scientific World Journal**, v. 2012, 2012.
- KIM, Kang-Chang et al. Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. **The Plant Cell**, v. 20, n. 9, p. 2357-2371, 2008.
- KIM, Shin-II et al. BRG1 requirement for long-range interaction of a locus control region with a downstream promoter. **Proceedings of the National Academy of Sciences**, v. 106, n. 7, p. 2259-2264, 2009.
- KIM, Tae-Houn et al. Guard cell signal transduction network: advances in understanding abscisic acid, CO<sub>2</sub>, and Ca<sup>2+</sup> signaling. **Annual review of plant biology**, v. 61, p. 561-591, 2010.
- KRZYWINSKI, Martin et al. Circos: an information aesthetic for comparative genomics. **Genome research**, v. 19, n. 9, p. 1639-1645, 2009.
- KLOSE, Robert J.; ZHANG, Yi. Regulation of histone methylation by demethylination and demethylation. **Nature reviews Molecular cell biology**, v. 8, n. 4, p. 307-318, 2007.
- KOMANDER, David. The emerging complexity of protein ubiquitination. **Biochemical Society Transactions**, v. 37, n. 5, p. 937-953, 2009.
- KOOYERS, Nicholas J. The evolution of drought escape and avoidance in natural herbaceous populations. **Plant Science**, v. 234, p. 155-162, 2015.
- KORASICK, David A. et al. Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. **Proceedings of the National Academy of Sciences**, v. 111, n. 14, p. 5427-5432, 2014.
- KUMAR, S. Vinod; WIGGE, Philip A. H2A. Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. **Cell**, v. 140, n. 1, p. 136-147, 2010.
- KUMAR, Sudhir; STECHER, Glen; TAMURA, Koichiro. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular biology and evolution**, v. 33, n. 7, p. 1870-1874, 2016.
- KUO, Min-Hao; ALLIS, C. David. Roles of histone acetyltransferases and deacetylases in gene regulation. **Bioessays**, v. 20, n. 8, p. 615-626, 1998.
- LALONDE, Jean-François; NARASIMHAN, Srinivasa G.; EFROS, Alexei A. What do the sun and the sky tell us about the camera?. **International Journal of Computer Vision**, v. 88, n. 1, p. 24-51, 2010.
- LECHNER, E. et al. MATH/BTB CRL3 receptors target the homeodomain-leucine zipper ATHB6 to modulate abscisic acid signaling. **Developmental cell**, v. 21, n. 6, p. 1116-1128, 2011.
- LEE, Joanne E.; GOLZ, John F. Diverse roles of Groucho/Tup1 co-repressors in plant growth and development. **Plant signaling & behavior**, v. 7, n. 1, p. 86-92, 2012.
- LESCOT, Magali et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. **Nucleic acids research**, v. 30, n. 1, p. 325-327, 2002.

- LIPMAN, David J.; WILBUR, W. John. Rapid similarity searches of nucleic acid and protein data banks. **Proceedings of the National Academy of Sciences**, v. 80, n. 3, p. 726-730, 1983.
- LIU, Chunyan et al. Histone methylation in higher plants. **Annual review of plant biology**, v. 61, p. 395-420, 2010.
- LOCKHART, David J.; WINZELER, Elizabeth A. Genomics, gene expression and DNA arrays. **Nature**, v. 405, n. 6788, p. 827-836, 2000.
- LUGER, Karolin; RICHMOND, Timothy J. The histone tails of the nucleosome. **Current opinion in genetics & development**, v. 8, n. 2, p. 140-146, 1998.
- LUO, Ming et al. Chromatin modifications and remodeling in plant abiotic stress responses. **Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms**, v. 1819, n. 2, p. 129-136, 2012.
- LUSSER, Alexandra; KÖLLE, Doris; LOIDL, Peter. Histone acetylation: lessons from the plant kingdom. **Trends in plant science**, v. 6, n. 2, p. 59-65, 2001.
- MALLIK, Rwitie; KUNDU, Anindya; CHAUDHURI, Shubho. High mobility group proteins: the multifaceted regulators of chromatin dynamics. **The Nucleus**, v. 61, n. 3, p. 213-226, 2018.
- MANNERS, John M.; CASU, Rosanne E. Transcriptome analysis and functional genomics of sugarcane. **Tropical Plant Biology**, v. 4, n. 1, p. 9-21, 2011.
- MARCHLER-BAUER, Aron et al. CDD: NCBI's conserved domain database. **Nucleic acids research**, v. 43, n. D1, p. D222-D226, 2015.
- MARTIN, Jeffrey A.; WANG, Zhong. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, n. 10, p. 671-682, 2011.
- MATHUR, Manisha. Bioinformatics challenges: A review. **Bioinformatics**, v. 3, n. 6, 2018.
- MATSUMURA, Hideo et al. SuperSAGE: a modern platform for genome-wide quantitative transcript profiling. **Current pharmaceutical biotechnology**, v. 9, n. 5, p. 368-374, 2008.
- MATSUMURA, Hideo et al. High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. **PloS one**, v. 5, n. 8, 2010.
- MILLER, Gad; SHULAEV, Vladimir; MITTLER, Ron. Reactive oxygen signaling and abiotic stress. **Physiologia plantarum**, v. 133, n. 3, p. 481-489, 2008.
- MOORE, Paul H.; BOTHA, Frederik C. (Ed.). **Sugarcane: physiology, biochemistry and functional biology**. John Wiley & Sons, 2013.
- MOLASIOTIS, Athanasios; FOTOPOULOS, V. **Oxidative and nitrosative signaling in plants**. Aristotle University of Thessaloniki, 2011.
- MORTAZAVI, A. Williams B a, McCue K, Schaeffer L, Wold B. **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. **Nat Methods**, v. 5, p. 621-628, 2008.

MUKHERJEE, Sunil Kumar. Origin and distribution of Saccharum. **Botanical Gazette**, v. 119, n. 1, p. 55-61, 1957.

NAKASHIMA, Kazuo; YAMAGUCHI-SHINOZAKI, Kazuko; SHINOZAKI, Kazuo. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. **Frontiers in plant science**, v. 5, p. 170, 2014.

NDAMUKONG, Ivan et al. Phosphatidylinositol 5-phosphate links dehydration stress to the activity of ARABIDOPSIS TRITHORAX-LIKE factor ATX1. **PLoS One**, v. 5, n. 10, 2010.

NEEDLEMAN, Saul B.; WUNSCH, Christian D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **Journal of molecular biology**, v. 48, n. 3, p. 443-453, 1970.

NOTREDAME, Cédric; HIGGINS, Desmond G.; HERINGA, Jaap. T-Coffee: A novel method for fast and accurate multiple sequence alignment. **Journal of molecular biology**, v. 302, n. 1, p. 205-217, 2000.

PAGE, Roderic DM. Tree View: An application to display phylogenetic trees on personal computers. **Bioinformatics**, v. 12, n. 4, p. 357-358, 1996.

PANDEY, Ritu et al. Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. **Nucleic acids research**, v. 30, n. 23, p. 5036-5055, 2002.

PATERSON, Andrew H. et al. The Sorghum bicolor genome and the diversification of grasses. **Nature**, v. 457, n. 7229, p. 551-556, 2009.

PEIRATS-LLOBET, Marta et al. A direct link between abscisic acid sensing and the chromatin-remodeling ATPase BRAHMA via core ABA signaling pathway components. **Molecular plant**, v. 9, n. 1, p. 136-147, 2016.

PETERSEN, Thomas Nordahl et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature methods**, v. 8, n. 10, p. 785, 2011.

PFAFFL, Michael W.; HORGAN, Graham W.; DEMPFLER, Leo. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic acids research**, v. 30, n. 9, p. e36-e36, 2002.

PINTO, Luciana Rossini et al. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. **Genome**, v. 47, n. 5, p. 795-804, 2004.

PIRRELLO, Julien et al. Functional analysis and binding affinity of tomato ethylene response factors provide insight on the molecular bases of plant differential responses to ethylene. **BMC Plant Biology**, v. 12, n. 1, p. 190, 2012.

POOLE, Rebecca L. The TAIR database. In: **Plant Bioinformatics**. Humana Press, 2005. p. 179-212.

- PRAKASHAM, R. S.; RAO, R. Sreenivas; HOBBS, Phil J. Current trends in biotechnological production of xylitol and future prospects. **Current Trends in Biotechnology and Pharmacy**, v. 3, n. 1, p. 8-36, 2009.
- RAE, Anne L. et al. Sucrose accumulation in the sugarcane stem: pathways and control points for transport and compartmentation. **Field Crops Research**, v. 92, n. 2-3, p. 159-168, 2005.
- RASHOTTE, Aaron M. et al. A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. **Proceedings of the National Academy of Sciences**, v. 103, n. 29, p. 11081-11085, 2006.
- RASMUSSEN, Randy. Quantification on the LightCycler. In: **Rapid cycle real-time PCR**. Springer, Berlin, Heidelberg, 2001. p. 21-34.
- REEVES, Raymond. HMG nuclear proteins: Linking chromatin structure to cellular phenotype. **Biochimica et biophysica acta**, v. 1799, n. 1-2, p. 3, 2010.
- RICHARDS, Richard A. Physiological traits used in the breeding of new cultivars for water-scarce environments. **Agricultural water management**, v. 80, n. 1-3, p. 197-211, 2006.
- Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA). Disponível em: <<http://www.ridesa.agro.ufg.br/>>, Acesso em: 07 jan, 2020.
- RIECHMANN, José Luis et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. **Science**, v. 290, n. 5499, p. 2105-2110, 2000.
- ROBERTSON, Neil et al. DiscoverySpace: an interactive data analysis application. **Genome biology**, v. 8, n. 1, p. R6, 2007.
- RODRIGUES, Fabiana Aparecida; DE LAIA, Marcelo Luiz; ZINGARETTI, Sonia Marli. Analysis of gene expression profiles under water stress in tolerant and sensitive sugarcane plants. **Plant Science**, v. 176, n. 2, p. 286-302, 2009.
- ROZEN, Steve; SKALETSKY, Helen. Primer3 on the WWW for general users and for biologist programmers. In: **Bioinformatics methods and protocols**. Humana Press, Totowa, NJ, 2000. p. 365-386.
- RYAN, Michael G. Tree responses to drought. **Tree Physiology**, v. 31, n. 3, p. 237-239, 2011.
- SAHA, S. Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE. Using the transcriptome to annotate the genome. **Nat Biotechnol**, v. 20, p. 508-512, 2002.
- SAHA, S. et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, v. 270, n. 5235, p. 467-470, 1995.
- SCARPARE, Fábio Vale et al. Sugarcane land use and water resources assessment in the expansion area in Brazil. **Journal of cleaner production**, v. 133, p. 1318-1327, 2016.
- SHAHBAZIAN, Mona D.; GRUNSTEIN, Michael. Functions of site-specific histone acetylation and deacetylation. **Annu. Rev. Biochem.**, v. 76, p. 75-100, 2007.

- SHEN, Yuan et al. Over-expression of histone H3K4 demethylase gene JMJ15 enhances salt tolerance in Arabidopsis. **Frontiers in plant science**, v. 5, p. 290, 2014.
- SHI, Yu-Jiang et al. Regulation of LSD1 histone demethylase activity by its associated factors. **Molecular cell**, v. 19, n. 6, p. 857-864, 2005.
- SHIMO, Hugo Massayoshi et al. Role of the Citrus sinensis RNA deadenylase CsCAF1 in citrus canker resistance. **Molecular plant pathology**, v. 20, n. 8, p. 1105-1118, 2019.
- SHIMKETS, Richard A. (Ed.). **Gene expression profiling**. Springer Science & Business Media, 2004.
- SHINOZAKI, Kazuo; YAMAGUCHI-SHINOZAKI, Kazuko. Gene networks involved in drought stress response and tolerance. **Journal of experimental botany**, v. 58, n. 2, p. 221-227, 2007.
- SHRESTHA, Barsha; GURAGAIN, Bhuwan; SRIDHAR, Vaniyambadi V. Involvement of co-repressor LUH and the adapter proteins SLK1 and SLK2 in the regulation of abiotic stress response genes in Arabidopsis. **BMC plant biology**, v. 14, n. 1, p. 54, 2014.
- SMITH, Temple F. et al. Identification of common molecular subsequences. **Journal of molecular biology**, v. 147, n. 1, p. 195-197, 1981.
- SONG, Chun-Peng; GALBRAITH, David W. AtSAP18, an orthologue of human SAP18, is involved in the regulation of salt stress and mediates transcriptional repression in Arabidopsis. **Plant molecular biology**, v. 60, n. 2, p. 241-257, 2006.
- SONG, Tao et al. OsJMJ703, a rice histone demethylase gene, plays key roles in plant development and responds to drought stress. **Plant Physiology and Biochemistry**, v. 132, p. 183-188, 2018.
- SORENG, Robert J. et al. A worldwide phylogenetic classification of the Poaceae (Gramineae). **Journal of Systematics and Evolution**, v. 53, n. 2, p. 117-137, 2015.
- SRIDHA, Sunandini; WU, Keqiang. Identification of AtHD2C as a novel regulator of abscisic acid responses in Arabidopsis. **The Plant Journal**, v. 46, n. 1, p. 124-133, 2006.
- SUNDARAVELPANDIAN, Kalaipandian; CHANDRIKA, Nulu Naga Prafulla; SCHMIDT, Wolfgang. PFT 1, a transcriptional M mediator complex subunit, controls root hair differentiation through reactive oxygen species (ROS) distribution in Arabidopsis. **New Phytologist**, v. 197, n. 1, p. 151-161, 2013.
- SOCCOL, Carlos Ricardo et al. Bioethanol from lignocelluloses: status and perspectives in Brazil. **Bioresource technology**, v. 101, n. 13, p. 4820-4825, 2010.
- TAIZ, L and ZEIGER, E. *Plant Physiology*. (4th), Sinauer Associates, Inc.; ISBN 978-0878938568, 2006
- TATUSOV, Roman L.; KOONIN, Eugene V.; LIPMAN, David J. A genomic perspective on protein families. **Science**, v. 278, n. 5338, p. 631-637, 1997.
- TARCZY-HORNOCH, Peter; MINIE, Mark. Bioinformatics challenges and opportunities. In: **Medical Informatics**. Springer, Boston, MA, 2005. p. 63-94.

TAVERNA, Sean D. et al. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. **Nature structural & molecular biology**, v. 14, n. 11, p. 1025-1040, 2007.

THARUN, Sundaresan; PARKER, Roy. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p–7p complex on deadenylated yeast mRNAs. **Molecular cell**, v. 8, n. 5, p. 1075-1083, 2001.

THIEBAUT, Flávia et al. Computational identification and analysis of novel sugarcane microRNAs. **BMC genomics**, v. 13, n. 1, p. 290, 2012.

THIEBAUT, Flavia et al. Differential sRNA regulation in leaves and roots of sugarcane under water depletion. **PLoS One**, v. 9, n. 4, 2014.

THIMM, Oliver et al. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. **The Plant Journal**, v. 37, n. 6, p. 914-939, 2004.

THOMPSON, Julie D. et al. A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. **PloS one**, v. 6, n. 3, 2011.

THU, Nguyen Binh Anh et al. Differential expression of two-component system–related drought-responsive genes in two contrasting drought-tolerant soybean cultivars DT51 and MTD720 under well-watered and drought conditions. **Plant molecular biology reporter**, v. 33, n. 5, p. 1599-1610, 2015.

TIAN, Feng et al. PlantRegMap: charting functional regulatory maps in plants. **Nucleic acids research**, v. 48, n. D1, p. D1104-D1113, 2020.

TIWARI, Shiv B. et al. AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. **The Plant Cell**, v. 13, n. 12, p. 2809-2822, 2001.

TRAPNELL, Cole; PACHTER, Lior; SALZBERG, Steven L. TopHat: discovering splice junctions with RNA-Seq. **Bioinformatics**, v. 25, n. 9, p. 1105-1111, 2009.

TREWICK, Sarah C.; MCLAUGHLIN, Paul J.; ALLSHIRE, Robin C. Methylation: lost in hydroxylation?. **EMBO reports**, v. 6, n. 4, p. 315-320, 2005.

TURNER, Bryan M. Cellular memory and the histone code. **Cell**, v. 111, n. 3, p. 285-291, 2002.

União da Indústria da cana-de-açúcar (UNICA). Disponível em: <https://www.unica.com.br/>. Acesso em: 08, Jan 2020.

VAN DIJK, Karin et al. Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in *Arabidopsis thaliana*. **BMC plant biology**, v. 10, n. 1, p. 238, 2010.

VELCULESCU, Victor E. et al. Serial analysis of gene expression. **Science**, v. 270, n. 5235, p. 484-487, 1995.

RAYA-GONZÁLEZ, Javier et al. PHYTOCHROME AND FLOWERING TIME1/MEDIATOR25 regulates lateral root formation via auxin signaling in *Arabidopsis*. **Plant physiology**, v. 165, n. 2, p. 880-894, 2014.

- WACLAWOVSKY, Alessandro J. et al. Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. **Plant Biotechnology Journal**, v. 8, n. 3, p. 263-276, 2010.
- WANG, Pengjie et al. Genome-wide identification of WOX genes and their expression patterns under different hormone and abiotic stress treatments in tea plant (*Camellia sinensis*). **Trees**, v. 33, n. 4, p. 1129-1142, 2019.
- WANG, Wenli et al. Genome-wide Analysis and expression divergence of the Trihelix family in *Brassica Rapa*: insight into the evolutionary patterns in plants. **Scientific reports**, v. 7, n. 1, p. 1-15, 2017.
- SMITH, Temple F. et al. Identification of common molecular subsequences. **Journal of molecular biology**, v. 147, n. 1, p. 195-197, 1981.
- WATERMAN, Michael S. Efficient sequence alignment algorithms. **Journal of Theoretical Biology**, v. 108, n. 3, p. 333-337, 1984.
- WEBER, Henriette et al. Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZ-MATH protein family. **Plant Physiology**, v. 137, n. 1, p. 83-93, 2005.
- WEBER, Henriette; HELLMANN, Hanjo. Arabidopsis thaliana BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. **The FEBS journal**, v. 276, n. 22, p. 6624-6635, 2009.
- WHITFIELD, Eleanor J.; PRUESS, Manuela; APWEILER, Rolf. Bioinformatics database infrastructure for biotechnology research. **Journal of biotechnology**, v. 124, n. 4, p. 629-639, 2006.
- WU, Shu et al. BRG1, the ATPase subunit of SWI/SNF chromatin remodeling complex, interacts with HDAC2 to modulate telomerase expression in human cancer cells. **Cell Cycle**, v. 13, n. 18, p. 2869-2878, 2014.
- YANAGISAWA, Shuichi. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. **Plant and Cell Physiology**, v. 45, n. 4, p. 386-391, 2004.
- YANG, Yan; LI, Ling; QU, Li-Jia. Plant Mediator complex and its critical functions in transcription regulation. **Journal of integrative plant biology**, v. 58, n. 2, p. 106-118, 2016.
- YILMAZ, Alper et al. GRASSIUS: a platform for comparative regulatory genomics across the grasses. **Plant physiology**, v. 149, n. 1, p. 171-180, 2009.
- YOLCU, Seher et al. Histone acetylation influences the transcriptional activation of POX in *Beta vulgaris* L. and *Beta maritima* L. under salt stress. **Plant Physiology and Biochemistry**, v. 100, p. 37-46, 2016.
- YU, Xiaoming et al. The chromatin remodeler ZmCHB101 impacts expression of osmotic stress-responsive genes in maize. **Plant molecular biology**, v. 97, n. 4-5, p. 451-465, 2018.
- ZINGARETTI, Sonia Marli et al. Sugarcane responses at water deficit conditions. **Water stress**, p. 255-276, 2012.

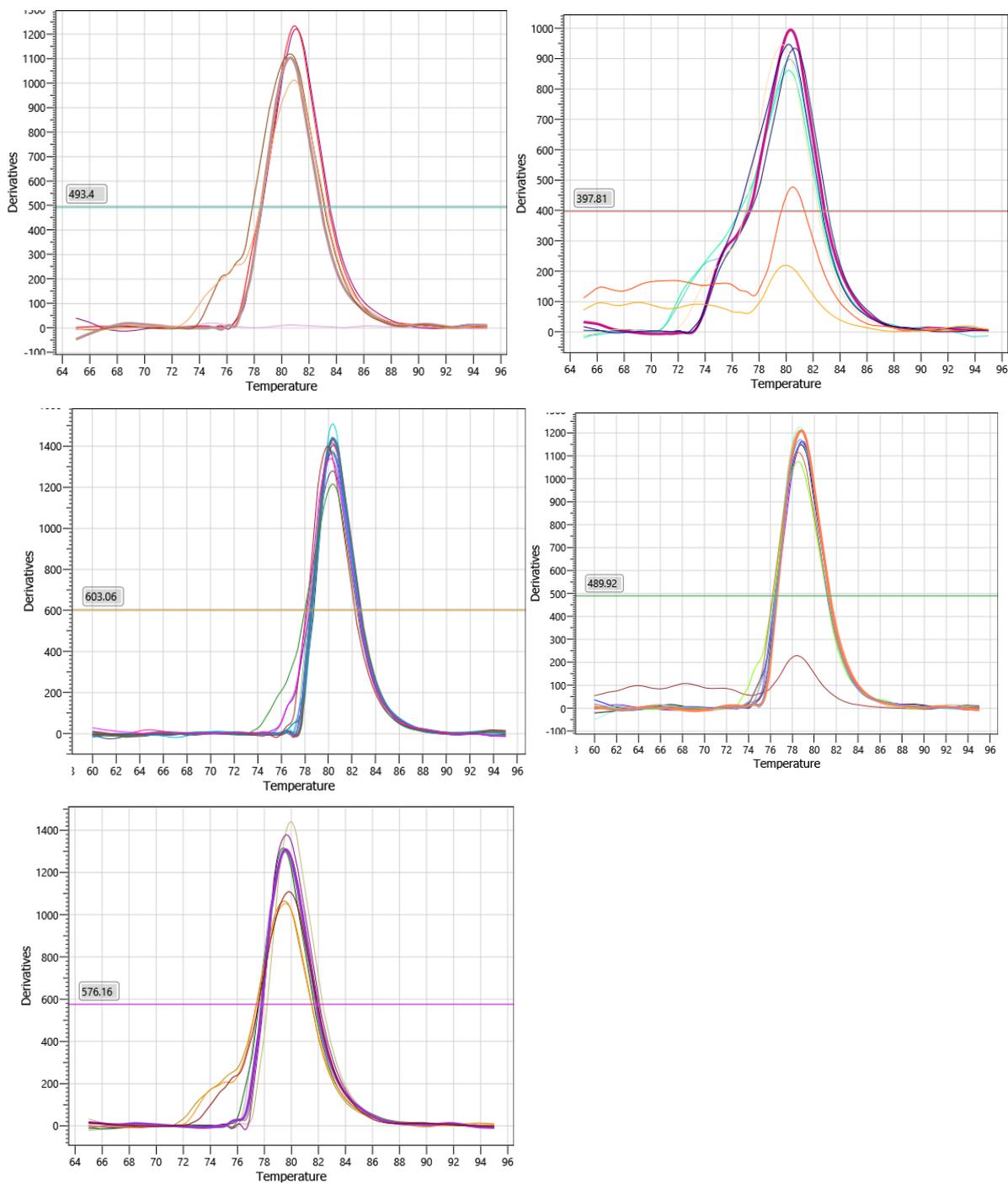
ZHANG, Zhaoliang et al. Arabidopsis floral initiator SKB1 confers high salt tolerance by regulating transcription and pre-mRNA splicing through altering histone H4R3 and small nuclear ribonucleoprotein LSM4 methylation. **The Plant Cell**, v. 23, n. 1, p. 396-411, 2011.

ZHENG, Yi et al. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. **Molecular plant**, v. 9, n. 12, p. 1667-1670, 2016.

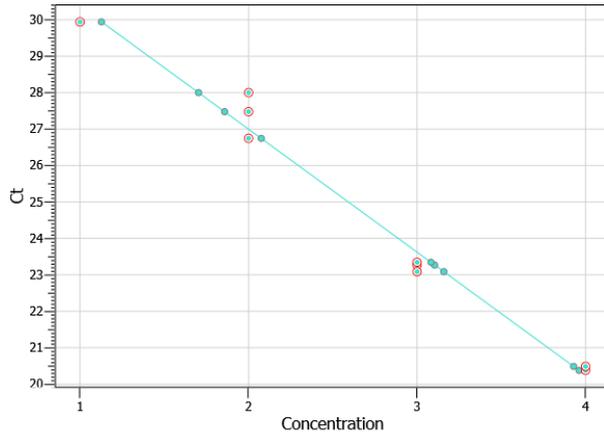
ZHU, Yongyou et al. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. **Molecular cell**, v. 49, n. 2, p. 298-309, 2013.

ZONG, Wei et al. Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. **Plant molecular biology**, v. 81, n. 1-2, p. 175-188, 2013.

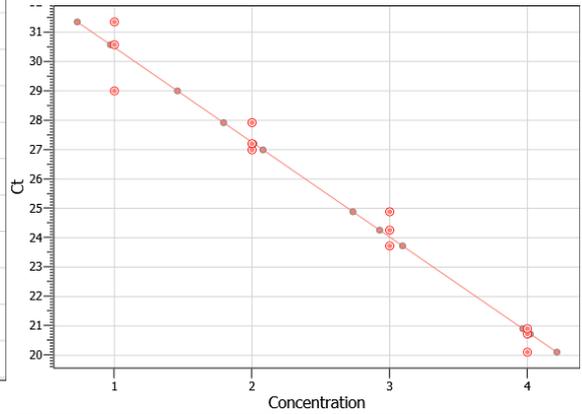
## ANEXOS



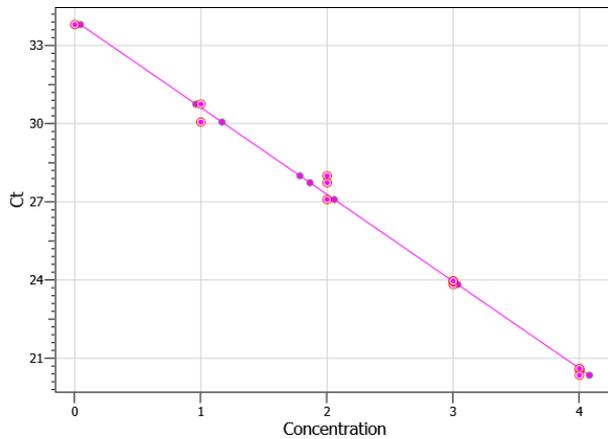
ANEXO A - CURVAS DE *MELTING* MOSTRANDO OS PICOS DE CADA ALVO AVALIADOS EM CDNAs DE RAÍZES DE ACESSOS DE CANA-DE-AÇÚCAR SUBMETIDOS A SUPRESSÃO DE REGA POR 24 H



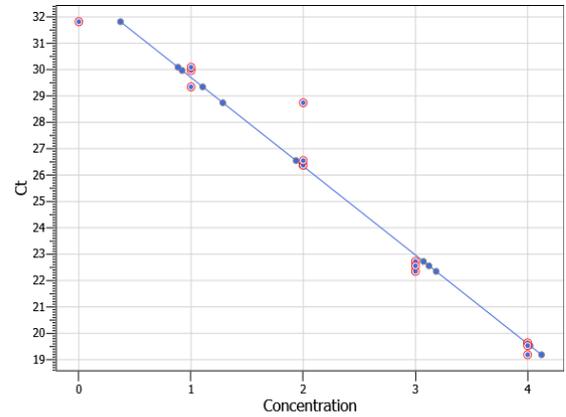
Assay:G9G10 - SYBR Intercept:33.75 Slope:-3.37 Error:0.019 Correlation:-0.989 Efficiency:97.84



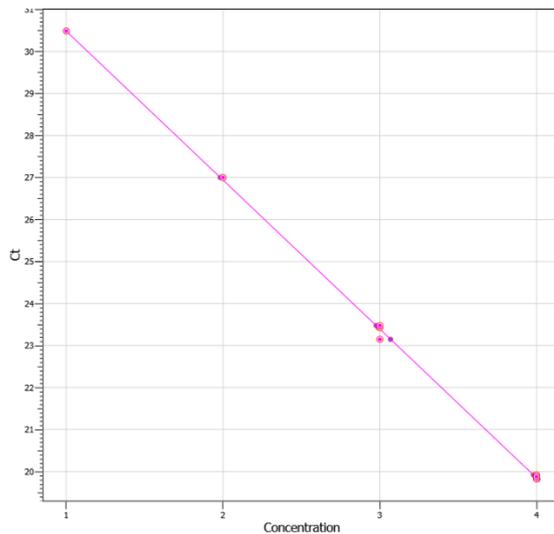
Assay:11A12A - SYBR Intercept:33.7 Slope:-3.23 Error:0.025 Correlation:-0.985 Efficiency:104.16



Assay:5B6B - SYBR Intercept:33.95 Slope:-3.33 Error:0.012 Correlation:-0.997 Efficiency:99.47



Assay:9A10A - SYBR Intercept:33.06 Slope:-3.37 Error:0.032 Correlation:-0.982 Efficiency:98.22



Assay:H1H2 - SYBR Intercept:34.63 Slope:-3.55 Error:0.004 Correlation:-1.000 Efficiency:91.46

ANEXO B - CURVAS PADRÃO (1:1, 1:10, 1:100, 1:1000, 1:10000) DE CADA ALVO AVALIADOS EM RAÍZES DE ACESSOS DE CANA-DE-AÇÚCAR SOB SUPRESSÃO DE REGA POR 24 H.