

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP**

**CÂMPUS DE JABOTICABAL**

**ANÁLISE DO TRANSCRIPTOMA DE FOLHAS DE CANA-DE-  
AÇÚCAR SUBMETIDAS À PROLONGADA LIMITAÇÃO  
HÍDRICA USANDO RNA-Seq**

**Aline Andruciole Belesini**

Bióloga

**2015**

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP**

**CÂMPUS DE JABOTICABAL**

**ANÁLISE DO TRANSCRIPTOMA DE FOLHAS DE CANA-DE-  
AÇÚCAR SUBMETIDAS À PROLONGADA LIMITAÇÃO  
HÍDRICA USANDO RNA-Seq**

**Aline Andruciole Belesini**

**Orientadora: Profa. Dra. Maria Inês Tiraboschi Ferro**

**Coorientador: Prof. Dr. Daniel Guariz Pinheiro**

**Coorientadora: Dra. Flávia Maria de Souza Carvalho**

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Agronomia (Genética e Melhoramento de Plantas).

**2015**

B428a Belesini, Aline Andrucioi  
Análise do transcriptoma de folhas de cana-de-açúcar submetidas à prolongada limitação hídrica usando RNA-Seq / Aline Andrucioi Belesini. -- Jaboticabal, 2015  
vii, 124 p. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2015  
Orientadora: Maria Inês Tiraboschi Ferro  
Coorientadores: Daniel Guariz Pinheiro, Flávia Maria de Souza Carvalho  
Banca examinadora: Poliana Fernanda Giachetto, Michael dos Santos Brito  
Bibliografia

1. Estresse hídrico. 2. Expressão gênica. 3. Montagem *de novo*. 4. *Saccharum spp.* 5. RNA-Seq. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 633.31



UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE JABOTICABAL

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

**TÍTULO: “ANÁLISE DO TRANSCRIPTOMA DE FOLHAS DE CANA-DE-AÇÚCAR SUBMETIDAS À PROLONGADA LIMITAÇÃO HÍDRICA USANDO RNA-Seq”**

**AUTORA: ALINE ANDRUCIOLI BELESINI**

**ORIENTADORA: Profa. Dra. MARIA INES TIRABOSCHI FERRO**

**CO-ORIENTADOR: Prof. Dr. DANIEL GUARIZ PINHEIRO**

**CO-ORIENTADORA: Profa. Dra. FLAVIA MARIA DE SOUZA CARVALHO**

Aprovada como parte das exigências para obtenção do Título de MESTRE EM AGRONOMIA (GENÉTICA E MELHORAMENTO DE PLANTAS), pela Comissão Examinadora:

Profa. Dra. MARIA INES TIRABOSCHI FERRO

Departamento de Tecnologia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Profa. Dra. POLIANA FERNANDA GIACHETTO

Empresa Brasileira de Pesquisa Agropecuária / Campinas/SP

Prof. Dr. MICHAEL DOS SANTOS BRITO

Instituto Agronômico de Campinas / Ribeirão Preto/SP

Data da realização: 03 de agosto de 2015.

## **DADOS CURRICULARES DA AUTORA**

**Aline Andruciole Belesini** – nascida em 24 de setembro de 1990, no município de Sertãozinho, São Paulo, filha de Denize Andruciole Belesini e Eduardo Aparecido Belesini. Formou-se em Ciências Biológicas pela Faculdade de Ciências Agrárias e Veterinárias – UNESP – Jaboticabal em 2012. Em março de 2013 iniciou o curso de mestrado junto ao Programa de Pós-Graduação em Agronomia (Genética e Melhoramento de Plantas) na Faculdade de Ciências Agrárias e Veterinárias (UNESP/Jaboticabal).

“Não existe triunfo sem perda, não há vitória sem sofrimento, não há liberdade sem sacrifício.”

**(J. R. R. Tolkien)**

Aos meus pais **Eduardo** e **Denize** por tudo que representam em minha vida. Pela confiança, incentivo e apoio ao longo dos anos. Amo vocês e sou grata por tudo que fizeram e fazem por mim!

Ao meu irmão **Arthur**, pela amizade, carinho, apoio e críticas durante esta jornada. Amo você!

Ao meu namorado **Samuel**, pelo amor, amizade e companheirismo em mais uma conquista. Obrigada por tudo, te amo!

**DEDICO COM AMOR**

## AGRADECIMENTOS

À Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Júlio de Mesquita Filho” (FCAV/UNESP) pela oportunidade de realização do mestrado;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos;

À Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) pela concessão da bolsa de estudos (Processo FAPESP n° 2013/22122-0);

À **Profa. Dra. Maria Inês T. Ferro**, pela orientação, ensinamentos, confiança, carinho, incentivo e amizade durante todo desenvolvimento desse trabalho.

Ao **Prof. Dr. Daniel Guariz Pinheiro**, pela coorientação e ajuda nas análises de bioinformática;

À **Dra. Flávia Maria de Souza Carvalho**, pela coorientação, ensinamentos, auxílios e preciosas dicas no desenvolvimento do projeto;

Ao **Prof. Dr. Jesus Aparecido Ferro**, pelos ensinamentos e disponibilidade do laboratório;

Ao **Prof. Dr. Jairo Osvaldo Cazetta** e toda sua equipe (Samira Cavallari, Bruna Telles e Thaís Silva) pela oportunidade de dar continuidade aos experimentos;

Aos componentes da banca de qualificação: **Prof. Dr. Daniel G. Pinheiro**, **Dra. Juliana da Silva Vantini** e **Dra. Samira D. C. Cavallari**, agradeço pela participação e contribuição.

Aos componentes da comissão examinadora: **Profa. Dra. Maria Inês T. Ferro**, **Dra. Poliana F. Giachetto** e **Dr. Michael dos Santos Brito** agradeço pela disponibilidade e participação.

À **Dra. Poliana Fernanda Giachetto** e **MSc. Giovanni Marques Castro**, pela disponibilidade, parceria e ajuda nas análises dos dados. Por toda paciência, conselhos e sugestões que enriqueceram o projeto, meu muito obrigado!

À **Dra. Juliana da Silva Vantini**, pela disponibilidade, correções, amizade e incentivo ao longo desses anos.

Ao **Prof. Dr. Luiz Lehmann Coutinho** e Ricardo pelo resequenciamento das amostras;



Aos queridos amigos do Laboratório de Bioquímica e Biologia Molecular: Rafael Marini, Fernando, Dani, Larissa, Jú Vantini, Flávia “mãe”, Julie, Tamiris, Aline Lopes, Thaísa, Angela, Amanda, Jéssica, Helen, Bruna, Ana, Clau e Flavinha; muito obrigada pelo maravilhoso convívio, pela amizade de cada um e por suas contribuições que foram de grande importância para a realização deste trabalho;

Às meninas do CREBIO: Agdinha, Má e Renata; agradeço o carinho, a disponibilidade e toda colaboração.

Às queridas Marina Gavassi e Thaís Angeli, que se mantiveram ao meu lado durante esses anos, com troca de experiências, incentivos e paciência. Sempre prontas a ouvir, com ótimos conselhos, e me dando todo apoio necessário. Agradeço demais por todos esses anos de amizade. Amo vocês!

À amiga-irmã, Bruna R. Telles, por compartilhar o projeto, as experiências, os momentos bons e ruins. Agradeço todo apoio e dedicação. Aprendi muito com você e não tenho palavras que consigam lhe agradecer o suficiente. Você é uma pessoa iluminada e muito especial! Este trabalho também é seu;

Às minhas gordinhas, Jana, Clau, Ana e Flavinha pela amizade e pelas contribuições. Vocês são muito queridas e vou levar um pouquinho de cada uma comigo ao longo da vida. Obrigada por tudo!

Ao colega Gustavo Claudiano, pelos momentos de descontração, pela preocupação e carinho;

A todos os funcionários, professores e colegas do Departamento de Tecnologia.

Aos meus amados pais, Denize e Eduardo, por não medirem esforços para que mais uma etapa fosse concluída. Mesmo com a distância sempre se fizeram presentes, dispostos a aconselhar e apoiar, foram e serão sempre meu suporte. Vocês são meus exemplos de vida, obrigada por me ensinarem tudo o que sou, amo vocês!

Ao meu irmão, Arthur, pela descontração e amizade. Pelas longas conversas e conselhos. Amo você!

Ao meu namorado Samuel, por todo amor, carinho e dedicação. Agradeço a paciência, companheirismo e o bom humor. Te amo!

Aos meus lindos cachorrinhos, Kinka e Fox, pela distração nos finais de semana, proporcionando sempre muita alegria e amor.

## SUMÁRIO

	Página
RESUMO.....	iii
ABSTRACT.....	iv
LISTA DE TABELAS .....	v
LISTA DE FIGURAS .....	vi
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA.....	4
2.1 Cana-de-açúcar: origem, aspectos gerais e importância econômica .....	4
2.2 Deficiência Hídrica .....	6
2.3 Transcriptoma .....	11
2.4 Tecnologias de Sequenciamento de Segunda Geração .....	13
2.5 RNA-seq.....	14
Figura 5. Etapa da clusterização na plataforma Illumina/Solexa. Fonte: Metzker (2010), citado por Cofre (2013).....	18
2.6 Montagem <i>de novo</i> .....	20
2.7 Estudos moleculares em cana-de-açúcar na resposta à seca .....	22
3 MATERIAL E MÉTODOS.....	25
3.1 Material vegetal e condições experimentais.....	25
3.2 Coleta das amostras .....	28
3.3 Extração de RNA .....	28
3.4 Isolamento do RNA mensageiro, montagem das bibliotecas de cDNA e ligação dos adaptadores.....	29
3.5 Sequenciamento do RNA.....	30
3.6 Clusterização das bibliotecas de cDNA e sequenciamento .....	31
3.7 Análise dos resultados do sequenciamento .....	31
3.7.1 Tratamento das leituras e filtro de qualidade .....	31
3.7.2 Montagem <i>de novo</i> do transcriptoma .....	32
3.7.3 Montagem de um transcriptoma referência (Clusterização) .....	32
3.7.4 Mapeamento e cálculo de abundância .....	33
3.7.5 Identificação dos Genes Diferencialmente Expressos (GDEs).....	33
3.7.6 Anotação dos transcritos .....	34
3.7.7 Integração dos dados e análise exploratória .....	35
3.7.8 Análise de agrupamentos de perfis de expressão gênica diferencial ...	35
4 RESULTADOS E DISCUSSÃO .....	36
4.1 Material Vegetal .....	36
4.2 Verificação da concentração e integridade do RNA total extraído .....	36
4.3 Análise do transcriptoma de folhas de cana-de-açúcar das duas cultivares aos 30, 60 e 90 dias de déficit hídrico .....	43
4.3.1 Tratamento das leituras e filtro de qualidade.....	43
4.3.2 Montagem <i>de novo</i> do transcriptoma .....	46

4.3.3	Identificação dos Genes Diferencialmente Expressos (GDEs).....	48
4.3.4	Anotação dos transcritos .....	51
4.3.5	Padrão geral da expressão diferencial das cultivares.....	55
4.3.6	Expressão gênica diferencial aos 30, 60 e 90 dias de estresse .....	68
5	CONCLUSÃO .....	105
6	REFERÊNCIAS.....	106

## ANÁLISE DO TRANSCRIPTOMA DE FOLHAS DE CANA-DE-AÇÚCAR SUBMETIDAS À PROLONGADA LIMITAÇÃO HÍDRICA USANDO RNA-Seq

**Resumo** - No Brasil, a cultura da cana-de-açúcar tem se expandido para regiões com prolongados períodos de deficiência hídrica, o que vem limitando a produção sucroenergética. Neste estudo, os perfis de expressão gênica de duas cultivares de cana-de-açúcar (*Saccharum* spp.), uma tolerante e outra sensível ao estresse hídrico, foram avaliados através da tecnologia de RNA-Seq. A montagem *de novo* do transcriptoma de folhas foi realizada visando identificar e analisar os genes diferencialmente expressos entre as cultivares, envolvidos com a resposta molecular durante períodos de estresse hídrico. Para isso, as duas cultivares foram plantadas em casa de vegetação e aos 60 dias após o plantio foram submetidas a três controlados potenciais hídricos do solo (controle, déficit hídrico moderado e déficit hídrico severo) e avaliadas aos 30, 60 e 90 dias após a aplicação dos tratamentos. A partir do método RNA-Seq, foram geradas mais de um bilhão de sequências, as quais permitiram obter um total de 177.509 e 185.153 sequências de transcritos, para as cultivares tolerante e sensível, respectivamente. O conjunto de transcritos expressos foram montados com o auxílio do programa Trinity. Estes transcritos foram alinhados com sequências de *Sorghum bicolor*, *Miscanthus giganteus*, *Arabidopsis thaliana* bem como sequências de cana-de-açúcar disponíveis em bancos públicos. Esta análise permitiu identificar o conjunto de genes de cana-de-açúcar compartilhados com outras espécies, bem como levou à identificação de novos transcritos ainda não catalogados. Os transcritos foram anotados e categorizados através dos termos do Gene Ontology (GO) em três categorias: componente celular, função molecular e processos biológicos para as duas cultivares. Os termos “regulador de enzima” e “regulador de transcrição” que se encontram dentro da categoria função molecular estão em evidência dentro dos termos associados aos genes diferencialmente expressos entre as cultivares contrastantes, sugerindo a importância da regulação gênica para as cultivares estudadas. Este estudo revelou diferentes padrões moleculares das cultivares envolvidas, o que proporcionou levantar hipóteses sobre a resposta da planta ao déficit hídrico e forneceu informações importantes quanto à identificação de genes envolvidos na resposta de tolerância à seca.

**Palavras-chave:** Estresse hídrico, Expressão gênica, Montagem *de novo*, *Saccharum* spp., RNA-Seq

## SUGARCANE LEAVES TRANSCRIPTOME ANALYSIS SUBMITTED TO PROLONGED WATER STRESS USING RNA-SEQ

**Abstract** – In Brazil, the sugarcane culture has expanded to areas with prolonged drought seasons, which is constraining sugarcane production. In this study, the gene expression profiles of two sugarcane cultivars (*Saccharum spp.*), one tolerant and other sensitive to water stress, were assessed by the RNA-Seq technology. The *de novo* assembly of leaves transcriptome was held aiming to identify and to analyze differentially expressed genes between the cultivars involved with molecular response during water stress periods. In order to accomplish this, the two cultivars were planted in a greenhouse and 60 days after planting them, they were submitted to three potential controlled water soil (control, moderate drought and severe water deficit) and evaluated at 30, 60 and 90 days after treatments application. Using the RNA-Seq method were generated over a billion sequences, which allowed to obtain a total of 177,509 and 185,153 transcripts sequences for the tolerant and sensitive cultivars, respectively. The set of expressed transcripts were assembled using the Trinity program. These transcripts were aligned with *Sorghum bicolor*, *Miscanthus giganteus*, *Arabidopsis thaliana* sequences and sugarcane sequences available in public databases. This analysis allowed the identification of the set of sugarcane genes shared with other species, as well as led to the identification of novel transcripts not cataloged yet. The transcripts were annotated and categorized by the terms of the Gene Ontology (GO) into three categories: cellular component, molecular function and biological process for the two cultivars. The terms "enzyme regulator" and "transcription regulator" that lie within the molecular function category are highlighted within the terms associated with the differentially expressed genes between the contrasting cultivars, suggesting the importance of gene regulation for the studied cultivars. This study found different molecular patterns of the involved cultivars, which provided hypotheses on plant response to drought and provided important information on the identification of genes involved in drought tolerance response.

**Keywords:** Drought stress, Gene expression, *De novo* Assembly, *Saccharum spp.*, RNA-Seq

## LISTA DE TABELAS

	Página
<b>Tabela 1.</b> Resultados obtidos com a leitura das amostras de RNA total nos três equipamentos: NanoDrop, Qubit e Bioanalyzer .....	37
<b>Tabela 2.</b> Resultado pré e pós-processamento inicial dos dados após o sequenciamento. ....	44
<b>Tabela 3.</b> Estatística descritiva da montagem dos transcriptomas das cultivares tolerante e sensível de cana-de-açúcar utilizando o programa Trinity. ....	47
<b>Tabela 4.</b> Número de genes diferencialmente expressos entre os tratamentos (linha x coluna). Os valores representam os genes diferencialmente expressos induzidos no tratamento representado na linha quando comparado ao da coluna. Valores acima da mediana estão em tons de verde e valores abaixo da mediana em tons de vermelho, valores próximos à mediana (3.028,5) estão em tons de amarelo. ....	49
<b>Tabela 5.</b> Resultados da anotação do transcriptoma da cana-de-açúcar contra os diferentes bancos de dados. ....	53
<b>Tabela 6.</b> Categorização dos genes diferencialmente expressos induzidos na cultivar tolerante (SP81-3250), em comparação à cultivar sensível (RB855453).....	59
<b>Tabela 7.</b> Categorização de genes mais expressos na cultivar sensível (RB855453), em comparação à cultivar tolerante (SP81-3250).....	64
<b>Tabela 8.</b> Descrição dos genes encontrados no Cluster 1 após blastx contra banco de dados não redundante (“nr”).....	76
<b>Tabela 9.</b> Descrição dos genes encontrados no Cluster 2 após blastx contra banco de dados não redundante (“nr”).....	86
<b>Tabela 10.</b> Descrição dos genes encontrados no Cluster 3 após blastx contra banco de dados não redundante (“nr”).....	93
<b>Tabela 11.</b> Descrição dos genes encontrados no Cluster 4 após blastx contra banco de dados não redundante (“nr”).....	100

## LISTA DE FIGURAS

	Página
<b>Figura 1.</b> Mapa de produção da cana-de-açúcar no Brasil. Fonte: UNICA, 2014.....	5
<b>Figura 2.</b> Mecanismo de percepção das plantas em resposta a diferentes estresses. Adaptado de BUCHANAN et al., 2000, citado por RIBEIRO, 2010. ....	8
<b>Figura 3.</b> Condição hídrica geral para o cultivo da cana-de-açúcar (safra 2014/2015) nos principais estados produtores do Brasil, no período de dezembro/2013 a novembro/2014. Fonte: Conab, 2014.....	10
<b>Figura 4.</b> Esquema geral de um experimento de expressão gênica pela técnica RNA-Seq na plataforma Illumina/Solexa. A etapa 1 engloba a preparação da amostra a partir do RNA extraído do organismo de estudo e sua conversão em uma biblioteca de cDNA. A etapa 2 é a clonagem in vitro para a amplificação da amostra, chamada de clusterização. A etapa 3 é o sequenciamento dos “clusters” formados de cada amostra. A etapa 4 compreende a análise dos resultados, com a filtragem dos dados, mapeamento das sequências contra um transcriptoma de referência, normalização dos dados e teste estatístico para aferir significância na diferença de expressão gênica de um determinado transcrito. A etapa 5 é o estudo biológico dos genes diferencialmente expressos, por meio da anotação gênica, categorização funcional dos genes e análises de vias metabólicas. Adaptado de JAIN (2011), citado por MURATA (2013). ....	17
<b>Figura 5.</b> Etapa da clusterização na plataforma Illumina/Solexa. Fonte: Metzker (2010), citado por Cofre (2013). ....	18
<b>Figura 6.</b> Representação esquemática do princípio tecnológico da plataforma Illumina. ....	19
<b>Figura 7.</b> Mini-toletes de cana-de-açúcar plantados em recipientes de plástico.....	26
<b>Figura 8.</b> Seleção e transplante das mudas para cada uma das unidades experimentais. ....	27
<b>Figura 9.</b> Aplicação das três distintas tensões hídricas em Mpa (T1- sem restrição hídrica, T2- restrição hídrica moderada, T3- restrição hídrica severa). A – Dez dias após aplicação do estresse; B – Sessenta dias após aplicação do estresse; C – Noventa dias após aplicação do estresse. ....	28
<b>Figura 10.</b> Equipamentos utilizados para a quantificação do RNA total. A- NanoDrop ND_1000; B- Qubit® 2.0 (Invitrogen); C- Bioanalyzer 2100 (Agilent);.....	29
<b>Figura 11.</b> Equipamentos utilizados para clusterização e sequenciamento. A - Flow cell, composta por oito canaletas, com a possibilidade de inserir até 12 amostras em cada uma delas ; B - cBot, equipamento utilizado para clusterização; C -Plataforma HiScanSQ.....	31
<b>Figura 12.</b> Eletroferograma e valor de RIN das amostras de RNA total de folhas de cana-de-açúcar coletadas aos 30 dias de estresse hídrico, gerados pelo Bioanalyzer (Agilent 2100 Bioanalyzer). ....	40
<b>Figura 13.</b> Eletroferograma e valor de RIN das amostras de RNA total de folhas de cana-de-açúcar coletadas aos 60 dias de estresse hídrico, gerados pelo Bioanalyzer (Agilent 2100 Bioanalyzer). ....	41

<b>Figura 14.</b> Eletroferograma e valor de RIN das amostras de RNA total de folhas de cana-de-açúcar coletadas aos 90 dias de estresse hídrico, gerados pelo Bioanalyzer (Agilent 2100 Bioanalyzer). .....	42
<b>Figura 15.</b> Eletroferograma e número de Integridade do RNA das amostras aos 30 dias de estresse gerados pelo Bioanalyzer. ....	43
<b>Figura 16.</b> Distribuição (%) das espécies mais representadas dentre as proteínas correspondentes aos genes da montagem do transcriptoma de cana-de-açúcar anotados com o banco de dados UniProtKB/Swiss-Prot.....	52
<b>Figura 17.</b> Categorização dos genes diferencialmente expressos comparando as duas cultivares, tolerante e sensível, submetidas ao déficit hídrico prolongado. Gráfico gerado pelo Programa WEGO, através da classificação fornecida pelo Gene Ontology. ....	56
<b>Figura 18.</b> Categorização dos genes diferencialmente expressos (GDEs) induzidos nas duas cultivares estudadas em cada um dos tempos amostrais. A- 30 dias; B-60 dias e C- 90 dias de déficit hídrico. ....	69
<b>Figura 19.</b> Diagramas de Venn mostrando o número de genes que foram induzidos nas cultivares tolerante (A) e sensível (B) nos tempos de 30, 60 e 90 dias após estresse. Os diagramas mostram o número de genes induzidos em cada tempo e o número de genes que são expressos em comum ou são exclusivos de cada tempo. ....	70
<b>Figura 20.</b> Comportamento de cinco genes induzidos na cultivar tolerante em relação à cultivar sensível (Cluster 1) ao longo do tempo (30, 60 e 90 dias).....	75
<b>Figura 21.</b> Regulação do estresse abiótico via sinalização por E3 ligases. A planta perceber os sinais de estresse através de sensores (desconhecido) e o sinal é enviado via hormônios vegetais, mensageiros secundários e reguladores da transcrição. A expressão induzida dos genes de estresse é facilitada por fatores de transcrição (TF). As E3 ligases tendem a regular os componentes da via de sinalização, principalmente os TFs. Na ausência de um sinal, as ligases podem suprimir a via de sinalização, por exemplo, promovendo a degradação de um TF. E3 ligases pode exercer uma função dentro de um mecanismo de feedback para melhorar ou atenuar o sinal do estresse. FONTE: LYZENGA; STONE, 2012. ....	83
<b>Figura 22.</b> Comportamento de quatro genes induzidos na cultivar tolerante em relação à cultivar sensível (Cluster 2) ao longo do tempo (30, 60 e 90 dias).....	85
<b>Figura 23.</b> Comportamento de dois genes induzidos na cultivar sensível em relação à cultivar tolerante (Cluster 3) ao longo do tempo (30, 60 e 90 dias). ....	92
<b>Figura 24.</b> Comportamento de dois genes induzidos na cultivar sensível em relação à cultivar tolerante (Cluster 4) ao longo do tempo (30, 60 e 90 dias). ....	99



## 1 INTRODUÇÃO

A cana-de-açúcar (*Saccharum* spp.) é uma cultura de grande importância mundial, devido a produção de açúcar e etanol, e também por se apresentar como uma das melhores opções para geração de bioenergia, por possuir elevada capacidade de produção de matéria seca (SILVA *et al.*, 2007).

Entretanto, nos últimos anos, esta cultura tem sido expandida para regiões do Brasil conhecidas por apresentarem longos períodos de restrição hídrica durante o ano, e os efeitos desse estresse sobre a produtividade da cultura, tais como diminuição nos teores de sacarose, baixo rendimento na produção, por exemplo, têm se tornando cada vez mais evidentes (CARLIN; SANTOS, 2009).

Sabe-se que a diminuição do teor de água no solo afeta alguns processos morfofisiológicos da cana-de-açúcar, causando assim distúrbios em seu metabolismo (PIMENTEL, 2004), e alterando as respostas fisiológicas e bioquímicas de acordo com o genótipo da planta, em intensa interação com o ambiente de produção (INMAN-BAMBER & SMITH, 2005).

Para contornar este problema e manter a produtividade vegetal sob condições de pouca umidade nos solos, há um grande interesse na identificação de características de ordem fisiológica, bioquímica e molecular durante a adaptação ao estresse que, posteriormente, poderão ser utilizadas na produção de cultivares tolerantes em programas de melhoramento genético (SILVA, 2010).

Na literatura é comum encontrar trabalhos com cana-de-açúcar em casa de vegetação envolvendo estresse hídrico drástico e aplicado em um curto período de tempo para a identificação de cultivares tolerantes (RODRIGUES *et al.*, 2009; RODRIGUES, F. A. *et al.*, 2011; GUPTA *et al.*, 2010) o que muitas vezes, não reflete os reais períodos que encontramos no campo e que afetam as principais etapas de desenvolvimento da planta.

As fases de perfilhamento e grande crescimento da cana-de-açúcar, também conhecidas como fases de formação (60 a 150 dias após o plantio), têm sido identificadas como as mais críticas por demanda de água (RAMESH, 2000; MACHADO *et al.*, 2009), principalmente porque são nestas fases onde

aproximadamente 70 a 80% da produção de cana é constituída (SINGH; RAO, 1987). Portanto, é neste período que experimentos de avaliação de cultivares nos programas de melhoramento, sob essas condições, deveriam ser conduzidos utilizando mecanismos apropriados, rápidos e eficientes (PINCELLI, 2010).

A tolerância ao estresse hídrico entre diferentes cultivares ou em estágios de desenvolvimento distinto das plantas pode resultar em alterações na expressão de genes envolvidos em diferentes etapas da resposta, desde a percepção do sinal até os mecanismos de transdução e regulação gênica (RODRIGUES et al., 2009; RODRIGUES, F.A. et al., 2011).

Um poderoso recurso para a análise das respostas das plantas ao estresse, é o estudo do transcriptoma (LESTARI et al., 2006). Uma das técnicas para tal estudo é o RNA-seq, que permite o mapeamento e a quantificação de transcritos (RODRIGUES, C.M., 2011), e segundo Ambrosone et al. (2013), essa ferramenta tem sido utilizada para descrever as respostas genéticas das plantas aos estímulos ambientais.

O método de RNA-seq permite uma detecção mais abrangente dos transcritos e suas isoformas, permitindo obter um perfil global dos genes expressos, além da expressão específica de um alelo, modificações pós-transcricionais ou com sequências rearranjadas (CARDOSO - SILVA et al., 2014).

Além disso, a detecção dos transcritos não fica restrita somente àqueles correspondentes a uma sequência genômica pré-existente, o que torna o método atrativo para a pesquisa em organismos cujos genomas ainda não foram determinados. Como não existem etapas de clonagem, o RNA-Seq requer uma quantidade muito menor de amostras de RNA e, uma vez que é um método quantitativo, pode ser usado para determinar os níveis de expressão gênica de maneira mais acurada. A técnica é capaz de capturar a dinâmica do transcriptoma através de diferentes tecidos e condições, o que é possível mesmo sem a aplicação de normalizações sofisticadas sobre os conjuntos de dados (WANG et al., 2009; OSHLACK et al., 2010). O método utiliza-se da clonagem *in vitro* e de sistemas de suporte sólido para as unidades de sequenciamento, o que permite que milhares de leituras possam ser determinadas de uma só vez com a plataforma (CARVALHO; SILVA, 2010; RODRIGUES, C. M., 2011).

Mesmo com sua importância econômica, o genoma da cana-de-açúcar ainda não está disponível, devido principalmente à sua complexidade genômica (alta ploidia) (KIM, et al., 2013). A base de recursos utilizada para o estudo das sequências de genes de cana atualmente são as ESTs (Expressed Sequence Tags), através de informações depositadas em bases de dados públicas (CARDOSO - SILVA et al., 2014), e o genoma do *Sorghum bicolor*, espécie filogeneticamente próxima da cana-de-açúcar, que já se encontra sequenciado (KIM et al., 2013).

Atualmente, todas as ESTs relatadas estão disponíveis no Sugarcane Gene Index, versão 3.0, que contém 282.683 ESTs e 499 sequências de cDNA completas, resultando em 121.342 sequências únicas montadas, ou unigenes. Porém, existe ainda mais de 10 mil genes a serem identificados, destacando - se a necessidade de novos esforços de sequenciamento do transcriptoma da cana (CARDOSO – SILVA et al., 2014).

Tendo em vista que a quantidade de informações disponíveis do genoma da cana ainda é limitada, uma vez que apenas algumas regiões foram sequenciadas e analisadas, uma alternativa aos estudos de expressão gênica, seria a montagem “*de novo*” do transcriptoma em questão, a fim de enriquecer o escasso conhecimento sobre o genoma desse organismo.

De acordo com o exposto acima, fica evidente que uma melhor compreensão e a correlação da expressão dos genes e seus produtos envolvidos nas respostas à deficiência hídrica prolongada se faz necessária para caracterizar os mecanismos que permitem a adaptação em condição limitante de água (RODRIGUES et al., 2009; GUPTA et al., 2010; ISKANDAR et al., 2011; RODRIGUES, F. A. et al., 2011), como os que se apresentam nas regiões do Brasil, onde a cana-de-açúcar tem se expandido.

Sendo assim, o objetivo do trabalho foi realizar a montagem *de novo* do transcriptoma de folhas da cana-de-açúcar e investigar, através da tecnologia de RNA-Seq, os transcritos diferencialmente expressos entre uma cultivar de cana-de-açúcar tolerante à seca e uma cultivar sensível, quando submetidas a três épocas amostrais (30, 60 e 90 dias após aplicação dos tratamentos).

## 2 REVISÃO DE LITERATURA

### 2.1 Cana-de-açúcar: origem, aspectos gerais e importância econômica

A cana-de-açúcar (*Saccharum* spp.) pertence à divisão Magnoliophyta, classe Liliopsida, ordem Cyperales, família Poaceae e gênero *Saccharum* (EMBRAPA, 2011). Essa família é conhecida como família das gramíneas, e inclui cereais como milho, trigo, arroz e sorgo, bem como algumas culturas forrageiras (PALHARES, 2010).

Provavelmente é originária do Sudeste Asiático, da grande região central da Nova Guiné e Indonésia (DANIELS; ROACH, 1987 *apud* SILVA, 2010). Da Nova Guiné, espalhou-se em várias linhas do sul do Oceano Pacífico, na Indochina, no Arquipélago da Malásia e em Bengala, aparecendo como planta produtora de açúcar na Índia (DELGADO; CÉSAR, 1977).

No Brasil, a cana-de-açúcar foi introduzida dois anos após o descobrimento do país, na Capitania de São Vicente, por Martin Afonso de Souza em 1502 (BARBOSA, 2007). Desde então, o Brasil se tornou o maior produtor mundial de cana-de-açúcar, e também o primeiro do mundo na produção de açúcar e etanol e conquista, cada vez mais, o mercado externo com o uso do biocombustível como alternativa energética (MAPA, 2015).

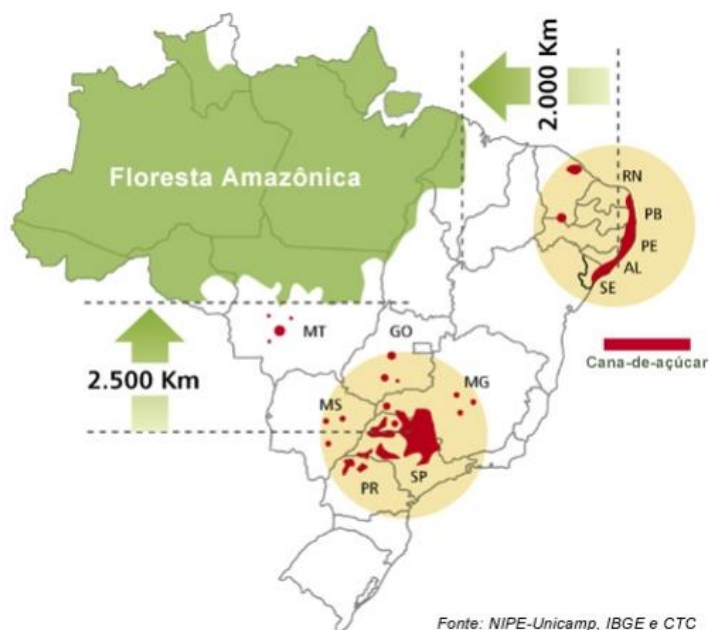
Trata-se de uma importante cultura para a produção de açúcar e etanol, sendo considerada uma das melhores opções para geração de bioenergia devido à sua elevada capacidade de produção de matéria seca (SILVA et al., 2007). A partir de 1975, com a criação do programa de incentivo à produção do álcool em larga escala, o Próalcool, a cultura da cana-de-açúcar ganhou um maior espaço e incentivo no cenário agrícola brasileiro (RIBEIRO, 2010).

Essa cultura mantém o maior sistema de produção de energia comercial de biomassa do mundo, através do etanol (substituindo cerca de 20% da gasolina) e da queima do bagaço (equivalente a 11 milhões de toneladas de óleo), sendo esta última, também uma fonte alternativa para a resolução da questão energética. Além de se destacar pelo aproveitamento racional da biomassa vegetal, gerando energia elétrica,

a cana-de-açúcar pode ser também utilizada na produção de papel e de plásticos biodegradáveis, de açúcares não calóricos, de compostos químicos de interesse farmacêutico, de produtos para a indústria etanol-química e como recurso forrageiro na alimentação suplementar dos ruminantes no período de estiagem. Neste sentido, a cana é hoje, uma das culturas que mais se destaca do ponto de vista econômico, social e ambiental (SACILOTO, 2003; MARTINS, 2004; DINIZ, 2007). No ano de 2012, o setor sucroalcooleiro foi responsável por aproximadamente 2% do PIB nacional e por 31% do PIB da agricultura no Brasil, tendo empregado cerca de 4,5 milhões de pessoas (PROCANA, 2012)

A produção de cana-de-açúcar se concentra nas regiões Centro-Sul e Nordeste do Brasil. A Figura 1, abaixo, mostra em vermelho as áreas onde se concentram as plantações e usinas produtoras de açúcar, etanol e bioeletricidade, segundo dados oficiais do IBGE, UNICAMP (Universidade Estadual de Campinas – SP) e do CTC (Centro de Tecnologia Canavieira) (ÚNICA, 2014).

A área destinada ao setor sucroalcooleiro brasileiro é atualmente de 9.004,5 mil hectares, que se distribuem pelos estados produtores de acordo com suas características. O estado de São Paulo, com 52% da produção e área de 4.685,7 mil hectares, junto aos estados de Goiás, Minas Gerais, Mato Grosso do Sul, Paraná, Alagoas e Pernambuco lideram a produção (CONAB, 2014).



**Figura 1.** Mapa de produção da cana-de-açúcar no Brasil. Fonte: UNICA, 2014.

O sucesso do cultivo da cana-de-açúcar se deve à utilização de variedades adaptadas aos diferentes ambientes edafoclimáticos, obtidas por meio de melhoramento genético clássico pelos centros de pesquisa e estações experimentais, que se trata de um processo relativamente demorado, uma vez que, dentre as espécies cultivadas, a cana-de-açúcar possui um dos genomas mais complexos, o que dificulta a aplicação de técnicas não convencionais de melhoramento (VETTORE et al., 2001).

Em 1999, foi criado o projeto SUCEST (“Sugarcane EST Genome Project”) que objetivou sequenciar genes transcritos em diferentes tecidos de cana-de-açúcar sob variadas condições fisiológicas (VETTORE et al., 2001). Os dados de sequências gerados por esse projeto, somados a utilização de ferramentas como bioinformática, macro e microarranjos, têm sido de primordial importância para a identificação de novos genes de cana-de-açúcar, bem como para estudos de genômica funcional, transcriptoma e expressão gênica (FERREIRA, 2008).

Dessa maneira, a biologia molecular assume papel-chave na identificação pontual de genes e de rotas metabólicas envolvidos nas respostas fisiológicas a estresses (CASAGRANDE et al., 2001) que variam de acordo com o genótipo da planta, em intensa interação com o ambiente de produção (INMAN-BAMBER; SMITH, 2005).

## **2.2 Deficiência Hídrica**

As plantas são submetidas frequentemente a condições adversas, resultando em estresses, que afetam de maneira desfavorável seu crescimento, desenvolvimento e/ou produtividade (GONÇALVES, 2008).

Esses estresses podem ser de origem biótica ou abiótica, podendo ocorrer de forma isolada ou concomitante. Dentre os abióticos, destacam-se: salinidade, déficit hídrico, deficiência ou excesso de nutrientes minerais, além de altas e baixas temperaturas (SOUZA, 2012).

Quando se trata de disponibilidade hídrica, a planta pode sofrer injúrias tanto por falta como por excesso de água, sendo que a primeira é mais comum na natureza, e exerce uma influência desvantajosa sobre a planta, causando desvio significativo

das condições ótimas para a vida, e induzindo à mudanças e respostas em todos os níveis funcionais do organismo (PINCELLI, 2010).

A deficiência hídrica pode ser definida como “todo o conteúdo de água de um tecido ou célula situado abaixo do conteúdo mais alto exibido no estado de maior hidratação” (TAIZ; ZEIGER, 2009).

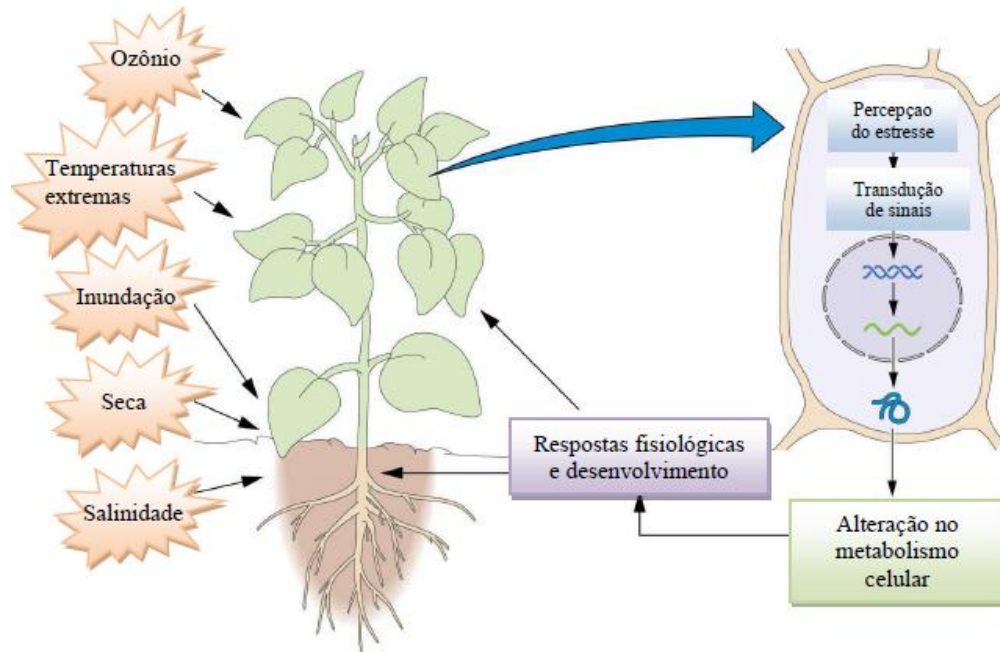
Segundo Vantini (2013), a seca pode reduzir significativamente os rendimentos das lavouras com implicações enormes, uma vez que produtores e toda a sociedade são afetados com o desemprego, aumento no preço de alimentos e instabilidade no mercado financeiro, por exemplo.

O déficit hídrico em plantas inicia-se a partir de uma complexa via de respostas, iniciando com a percepção da falta de água, que desencadeia uma sequência de eventos de ordem molecular, finalizando com várias respostas fisiológicas, metabólica e de desenvolvimento (BRAY, 1997).

As respostas das plantas a esse estresse vão depender de uma série de fatores, tais como: espécie, genótipo, duração e severidade, idade e estágio de desenvolvimento, órgão e tipo de célula e do comportamento sub-celular. Dessa forma, as plantas podem ser dotadas de mecanismos de tolerância ou resistência e suscetibilidade em seu crescimento, podendo chegar à morte, dependendo da severidade do estresse a que as mesmas são submetidas (CHAVES et al., 2003; CAMBRAIA, 2005).

Como proteção, as plantas tendem a diminuir a perda de água pelo fechamento parcial dos estômatos, causando acentuada senescência foliar e restrição ao surgimento de novas folhas (MACHADO et al., 2009). Além disso, ocorre também o alongamento radicular em direção às zonas do solo mais profundas, o que possibilita permanecerem úmidas por mais tempo (TAIZ; ZEIGER, 2009).

Pesquisas genéticas desenvolvidas a partir de diversas espécies vegetais, inclusive cana-de-açúcar, têm mostrado que as alterações fisiológicas e metabólicas das plantas em resposta ao déficit hídrico para uma maior tolerância a estresses ambientais ocorrem de acordo com os genes que possuem, e principalmente de que maneira estes genes são ativados (BUCHANAN et al., 2000; ARAÚJO et al., 2005) (Figura 2).



**Figura 2.** Mecanismo de percepção das plantas em resposta a diferentes estresses. Adaptado de BUCHANAN et al., 2000, citado por RIBEIRO, 2010.

No Brasil, a maior parte da cana-de-açúcar é cultivada em regime de sequeiro. Em virtude disso, a cultura está sujeita a ocorrência de déficit hídrico em determinados estágios do seu desenvolvimento (LYRA et al., 2012). Além disso, nos últimos anos a expansão do cultivo da cana-de-açúcar no Brasil avançou para as regiões do Oeste do Estado de São Paulo e áreas de cerrado, conhecidas por apresentarem longos períodos de restrição hídrica durante o ano, e os efeitos desse estresse sobre a produtividade da cultura têm se tornando cada vez mais evidentes (CARLIN; SANTOS, 2009).

Segundo dados da CONAB, a safra de 2014/2015 sofreu com as adversidades climáticas ocorridas nas lavouras de cana-de-açúcar durante o período de desenvolvimento, o que impactou diretamente as produtividades esperadas, que ficaram aquém das obtidas na safra passada. O decréscimo se concentrou com maior intensidade na Região Sudeste, onde as precipitações pluviométricas ficaram abaixo do normal e refletiu no desenvolvimento e rendimento da cultura. A queda no rendimento agrícola foi de 9,6%, chegando a atingir 10,5% em São Paulo (CONAB, 2014). Além disso, segundo dados da EMBRAPA (2014), o Brasil encontra-se em uma situação de seca inédita em função de dois fenômenos independentes. Primeiro,



devido a uma variação no clima da Terra, provocada por um ciclo natural no Oceano Pacífico e no Atlântico, que está retirando energia da atmosfera e absorvendo-a em suas águas profundas. No Brasil, isso reduz a frequência e a força das frentes frias que vêm do Polo Sul para o Sudeste e gerariam as chuvas que alimentariam as represas. O outro fenômeno é representado por uma zona de alta pressão que se forma sobre o Atlântico e impede a entrada das frentes frias. Elas só vão até o Paraná e não chegam a São Paulo, é o que se chama de bloqueio atmosférico.

Essa é uma condição normal no inverno, mas em 2013 o bloqueio não se dispersou depois dessa estação, o que impediu a entrada de frentes frias no verão e durou até outubro de 2014. Depois, as chuvas começaram a se normalizar, mas outro bloqueio se formou no meio de dezembro de 2014 e só começou a dispersar em janeiro/fevereiro de 2015. Isso já tinha sido observado antes, mas não com tamanha intensidade (EMBRAPA, 2014).

A ocorrência de chuvas abaixo da média histórica em algumas regiões do Brasil exige atenção e cuidadoso acompanhamento da produção e do abastecimento de alimentos no decorrer do ano. Aliada à baixa precipitação, as elevadas temperaturas intensificam a evapotranspiração, ocasionando baixa disponibilidade de água para as culturas. Na Figura 3, verificam-se as condições hídricas gerais nos principais estados produtores de cana-de-açúcar do Brasil (CONAB, 2014).



**Figura 3.** Condição hídrica geral para o cultivo da cana-de-açúcar (safra 2014/2015) nos principais estados produtores do Brasil, no período de dezembro/2013 a novembro/2014. Fonte: Conab, 2014.

A manutenção da produtividade vegetal sob condições de pouca umidade nos solos leva pesquisadores à identificação de características de ordem fisiológica, bioquímica e molecular durante a adaptação ao estresse que, posteriormente, serão utilizadas na produção de cultivares tolerantes, em programas de melhoramento genético (CARLIN; SANTOS, 2009; SILVA, 2010).

O estresse por seca influencia de maneira direta o crescimento dos perfilhos, a altura final dos colmos e, conseqüentemente, a produção de açúcar (SUGIHARTO, 2004). A identificação de características morfofisiológicas, bioquímicas e moleculares, observadas durante o processo de adaptação das plantas aos estresses, tem auxiliado na produção de plantas mais tolerantes, seja através do melhoramento genético convencional ou da engenharia genética (TESTER; BACIC, 2005; CHERIAN et al., 2006; BHATNAGAR-MATHUR et al., 2008).

Desde 2004, membros da RIDESA (Rede Interuniversitária para o Desenvolvimento Sucroalcooleiro), vêm conduzindo experimentos sob condições de campo e de casa-de-vegetação para estudar os mecanismos de tolerância à seca em cana-de-açúcar. Foram realizadas análises agrônômicas, fisiológicas e bioquímicas da cultura, e os resultados obtidos até o momento mostram que as variedades de cana-de-açúcar possuem mecanismos distintos de tolerância à seca (SOUZA, 2012).

Com relação às análises moleculares, a base de recursos utilizado para o estudo das sequências de genes de cana atualmente são as ESTs (Expressed Sequence Tags), através de informações depositadas em bases de dados públicas (CARDOSO - SILVA et al., 2014), e o genoma do *Sorghum bicolor*, organismo filogeneticamente mais próximo da cana-de-açúcar, que já se encontra sequenciado (KIM et al., 2013).

Atualmente, todas as ESTs relatadas estão disponíveis no Sugarcane Gene Index, versão 3.0, que contém 282.683 ESTs e 499 sequências de cDNA completas, resultando em 121.342 sequências únicas montadas, ou unigenes. Porém, existe ainda mais de 10 mil genes a serem identificados, destacando a necessidade de novos esforços para o sequenciamento da cana-de-açúcar (CARDOSO – SILVA et al., 2014).

Dessa forma, fica evidente que uma melhor compreensão e a correlação da expressão dos genes e produtos de sua expressão envolvidos nas respostas à deficiência hídrica prolongada se faz necessária para caracterizar os mecanismos que permitem a adaptação em condição limitante de água (RODRIGUES et al., 2009; GUPTA et al., 2010; ISKANDAR et al., 2011; RODRIGUES, F. A. et al., 2011).

### **2.3 Transcriptoma**

Desde o evento marcante da descoberta da estrutura dupla hélice do DNA, que permitiu um grande avanço na compreensão dos fenômenos que ocorrem nos organismos vivos e abriu a possibilidade de uma nova estratégia para a genética, a era genômica tornou-se uma realidade. Novas estratégias de pesquisa e tecnologias têm gerado um fluxo constante de dados genômicos cada vez maiores e mais complexos que são depositados em bancos de dados públicos e transformaram o estudo de praticamente todos os processos da vida (COLLINS et al., 2003).

Entretanto, o conhecimento da sequência e da estrutura de um gene não permite que se alcance o objetivo máximo desta pesquisa, que é a determinação da função de cada gene na célula (PASSOS; JORDAN, 2000; MARQUES; SILVA, 2004).

Apenas uma fração dos milhares de genes presentes no organismo é expressa em determinado momento. Para que a mensagem contida no DNA possa ser convertida em uma função celular, é necessário que os genes sejam transcritos em

outro tipo de molécula, um RNA, que posteriormente será traduzido em proteínas (NELSON; COX, 2004).

Acredita-se que o RNA celular é constituído principalmente de RNA ribossomal (rRNA 80 – 90%), RNA transportador ( tRNA 5-15%), RNA mensageiro (mRNA 2-4%) e RNA não codificante (1%, ncRNA) com funções de regulação indefinidos (COSTA, 2010).

A regulação da transcrição envolve não apenas as proporções diferentes da transcrição nas diferentes partes do genoma, como também a escolha das regiões que devem ser transcritas, e a extensão desta transcrição. Desse modo, diferentes conjuntos de genes podem ser transcritos em diferentes células, ou na mesma célula, em momentos diferentes (WATSON; BAKER, et al., 2006).

Dessa forma, o RNA foi considerado uma “ponte” de transferência de informações biológicas entre o DNA e as proteínas (COSTA, 2010). Criou-se o termo transcriptoma, que representa o conjunto completo dos transcritos (RNA) da célula, em um estágio específico de desenvolvimento ou condição fisiológica (NOBUTA, VENU, et al., 2007), que é chamado de transcriptoma celular (NELSON; COX, 2004).

Sendo assim, descobrir e interpretar a complexidade de um transcriptoma representa um objetivo crucial para a compreensão dos elementos funcionais de um genoma, revelando a complexidade do código genético de organismos vivos pela análise dos constituintes moleculares de células e tecidos (WANG; GERSTEIN; SNYDER, 2009).

O estudo do transcriptoma determina quando e onde os genes são ligados ou desligados em vários tipos de células e tecidos. Estes transcritos podem ser quantificados e revelar a expressão de determinados genes em uma célula. Além disso, com a análise do transcriptoma é possível gerar uma imagem completa de genes que são ativados em vários estágios de desenvolvimento de um organismo (ADAMS, 2008).

Vários estudos de transcriptoma têm sido realizados com cana-de-açúcar para elucidar seus mecanismos de defesa à diversos estresses bióticos e abióticos, tais como estresse salino (FERREIRA, 2008), pragas (MORAES, 2010), deficiência hídrica (CARDOSO - SILVA et al., 2014; VANTINI et al., 2015), dentre outros.

No estudo de transcriptomas podem ser utilizadas diferentes tecnologias: a) macro e microarranjos, baseadas na hibridização de sondas; b) sequenciamento de regiões específicas de fragmentos de cDNA; e c) análise de fragmentos de cDNA amplificados via PCR (WANG et al., 2009).

Algumas limitações na técnica de microarranjos, como a disponibilidade de um banco de dados sólido e a ocorrência de hibridação cruzada, dificultam a análise da expressão gênica. As técnicas de sequenciamento de regiões específicas de fragmentos de cDNA também possuem desvantagens, pois há necessidade de grande quantidade de mRNA, e apenas uma parte dos transcritos é analisada e geralmente suas isoformas são indistinguíveis umas das outras (WANG et al. 2009).

A partir de então, várias tecnologias foram aprimoradas para a aplicação em grande escala e na última década, houve um enorme avanço no desenvolvimento de métodos de dedução e quantificação dos níveis de expressão gênica ao nível do transcriptoma completo, tornando possível a análise simultânea da expressão de milhares de genes (WANG et al., 2009).

Novas tecnologias de sequenciamento de DNA em larga escala, denominadas tecnologias de sequenciamento de segunda geração, promovem o sequenciamento de DNA em plataformas capazes de gerar informação sobre milhões de fragmentos em uma única corrida. Sequenciadores de nova geração e suas plataformas possibilitaram um novo método de sequenciamento de bibliotecas de cDNA, denominado RNA-seq (HAAS, 2012), a qual permite o mapeamento e quantificação de transcriptomas (RODRIGUES, C. M., 2011), e que vem substituindo rapidamente as antigas abordagens no estudo de expressão gênica (WANG et al., 2009; FU et al., 2009; VAN VELLETT et al., 2010).

## **2.4 Tecnologias de Sequenciamento de Segunda Geração**

Nos últimos 30 anos, a biologia molecular foi submetida a grandes avanços e 2004 será lembrado como o ano que revolucionou o campo graças à introdução de plataformas de sequenciamento de nova geração (MARDIS, 2008; SHENDURE; JI, 2008; METZKER, 2010).

Dentre as novas plataformas de sequenciamento, as principais são: o pirosequenciamento da Roche/454, sequenciamento por síntese da Illumina/Solexa e sequenciamento por ligação da Life Technologies/SOLiD (WANG et al. 2009; JAIN, 2011).

As três principais plataformas possuem diferenças quanto ao tamanho de leitura, e o método de detecção, mas as tecnologias de sequenciamento são basicamente divididas nas seguintes etapas: preparação da amostra, sequenciamento, captura da imagem e a análise dos dados (METZKER, 2010).

No sequenciamento de segunda geração, o DNA genômico (ou cDNA) é quebrado aleatoriamente em tamanhos menores que, são ligados a uma superfície sólida ou indiretamente imobilizadas (por ligação de uma polimerase ao suporte). Essa imobilização permite que milhares de reações de sequenciamento sejam realizadas simultaneamente, gerando assim uma grande quantidade de leituras. A grande quantidade de dados gerados é muito atrativa, mas a eliminação do passo da clonagem dos fragmentos de DNA (ou cDNA) é o maior benefício destas novas tecnologias. Todos os métodos atuais permitem o uso direto de pequenos DNAs ou fragmentos de RNA sem sua inserção dentro de um plasmídeo ou de outro vetor, removendo assim uma etapa dispendiosa que era necessária para o sequenciamento tradicional do método de Sanger (COSTA, 2010).

As mais amplas aplicações estão se utilizando dessas novas tecnologias. Na verdade, essas plataformas foram rapidamente aplicadas a muitos contextos genômicos utilizando a terminação “seq” (por exemplo, RNA-Seq para transcriptoma, Chip-Seq para interação DNA-proteína, etc.) (COSTA, 2010).

## **2.5 RNA-seq**

Sequenciadores de nova geração e suas plataformas possibilitaram um novo método de sequenciamento de bibliotecas de cDNA, denominado RNA-seq (HAAS, 2012). Essa técnica é considerada pelos pesquisadores um método revolucionário e segundo Ambrosone et al. (2013), essa tecnologia tem sido amplamente utilizada para descrever as respostas genéticas das plantas aos estímulos ambientais.

O método de RNA-seq permite uma detecção mais abrangente dos transcritos e suas isoformas, permitindo obter um perfil global dos genes expressos, além da expressão específica de um alelo, modificações pós-transcricionais ou com sequências rearranjadas (CARDOSO - SILVA et al., 2014).

Além disso, a detecção dos transcritos não fica restrita somente àqueles correspondentes a uma sequência genômica pré-existente, o que torna o método atrativo para a pesquisa em organismos cujos genomas ainda não foram determinados. Como não existem etapas de clonagem, o RNA-Seq requer uma quantidade muito menor de amostras de RNA e, uma vez que é um método quantitativo, pode ser usado para determinar os níveis de expressão gênica de maneira mais acurada. A técnica é capaz de capturar a dinâmica do transcriptoma através de diferentes tecidos e condições, sem necessidade de que normalizações sofisticadas sejam aplicadas sobre os conjuntos de dados (WANG et al., 2009; OSHLACK et al., 2010).

O método utiliza-se da clonagem *in vitro* e de sistemas de suporte sólido para as unidades de sequenciamento, o que permite que milhares de leituras possam ser determinadas de uma só vez (CARVALHO & SILVA, 2010; RODRIGUES, C. M., 2011). O sequenciamento dos fragmentos no protocolo de RNA-Seq da Illumina/Solexa, tecnologia utilizada nesse estudo, pode ser *single-read* ou *paired-end*, ou seja, uma população de mRNA é convertida em uma biblioteca de cDNA fragmentada ao acaso com adaptadores ligados a uma (*single-read*) ou em ambas (*paired-end*) as extremidades. Em seguida, ocorre uma amplificação por PCR a partir de uma ou de ambas as extremidades (JAIN, 2011).

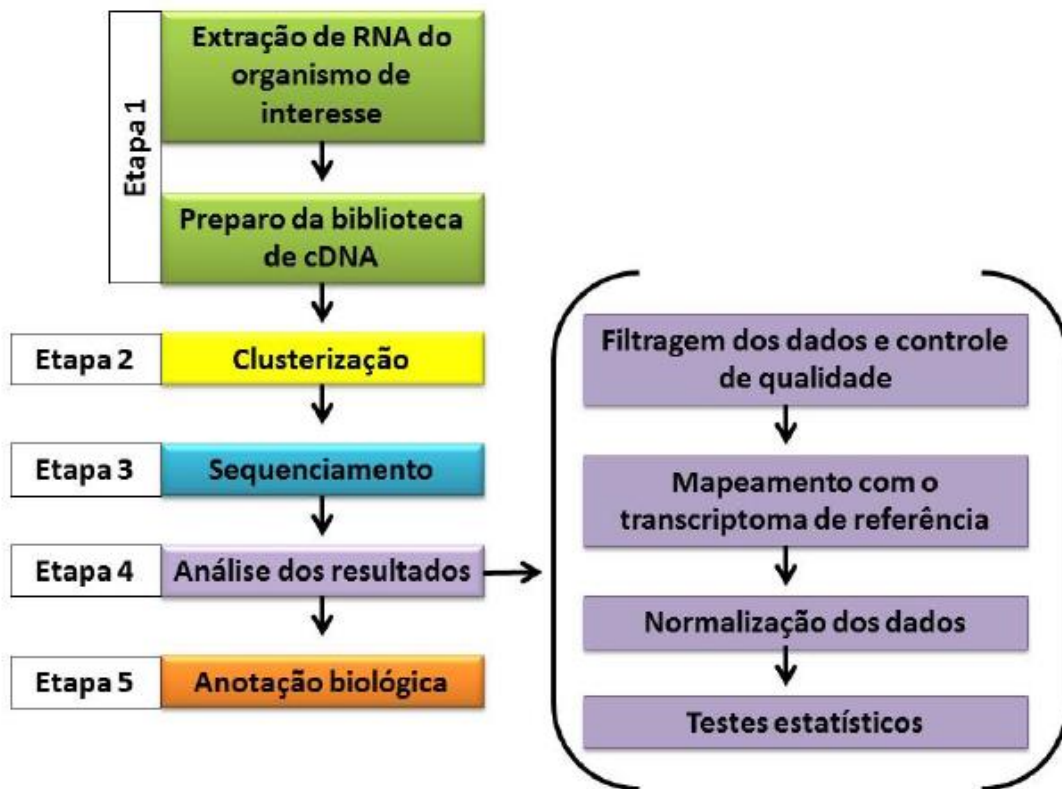
O processo é dividido em quatro etapas: 1) preparo das bibliotecas de cDNA; 2) clusterização; 3) sequenciamento e captura de imagem e 4) análise dos resultados (Figura 4). O estudo do transcriptoma, em geral, inicia-se com a extração de RNA total de boa qualidade do tecido e organismo de interesse. Para organismos eucariotos, o RNA mensageiro (mRNA) é isolado por meio de colunas de oligo dT e o restante dos RNAs, como o RNA ribossomal (rRNA), é eluído. No caso de organismos procariotos, é feita a remoção do rRNA da amostra com esferas magnéticas que contêm sondas que capturam este rRNA. Posteriormente, a molécula de mRNA é convertida na primeira fita de DNA complementar (cDNA) por meio da enzima transcriptase reversa.

Em seguida, a fita simples é convertida em uma molécula de cDNA de fita dupla pela enzima DNA polimerase. As moléculas de cDNA são fragmentadas aleatoriamente em fragmentos de tamanhos curtos pela ação da enzima DNase I. As extremidades destes fragmentos de cDNA são convertidas para extremidades cegas ou abruptas pela enzima T4 DNA polimerase. Em seguida, as extremidades são adeniladas para facilitar a ligação dos adaptadores. Posteriormente, os adaptadores são adicionados. Por fim, os fragmentos de cDNA são amplificados por uma reação de PCR seletiva, na qual apenas os fragmentos que contém adaptadores em ambas as extremidades são amplificados. Após esta etapa, as bibliotecas de cDNA estão prontas para o processo de clusterização.

Na clusterização, também conhecida como etapa de clonagem *in vitro*, utiliza-se uma superfície de clonagem denominada *flow cell*, que é dividida em oito linhas ou canaletas que são utilizadas para o sequenciamento das amostras. Em cada linha, adaptadores são fixados à superfície pela extremidade 5', deixando a extremidade 3' livre para servir na iniciação da reação de sequenciamento dos fragmentos imobilizados no suporte por hibridização (Figura 5).

Os fragmentos de cDNA da amostra também são ligados aos adaptadores em ambas as extremidades, o que permite sua fixação ao suporte de sequenciamento por hibridização a um dos adaptadores fixados.

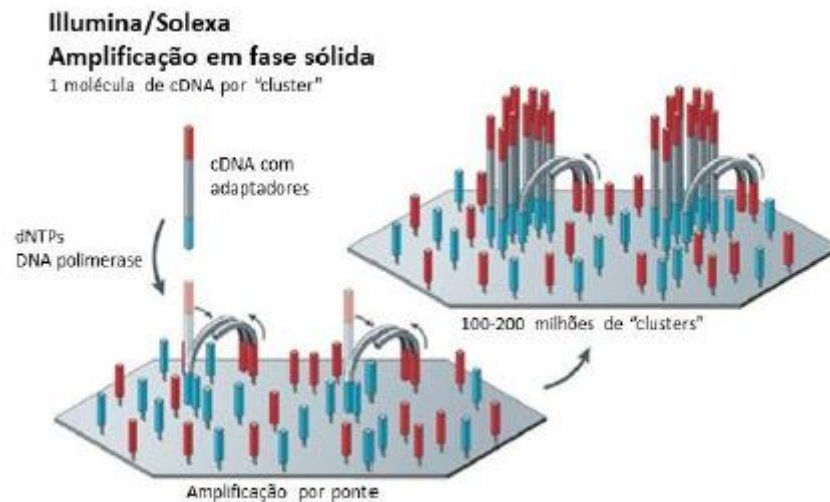




**Figura 4.** Esquema geral de um experimento de expressão gênica pela técnica RNA-Seq na plataforma Illumina/Solexa. A etapa 1 engloba a preparação da amostra a partir do RNA extraído do organismo de estudo e sua conversão em uma biblioteca de cDNA. A etapa 2 é a clonagem *in vitro* para a amplificação da amostra, chamada de clusterização. A etapa 3 é o sequenciamento dos “clusters” formados de cada amostra. A etapa 4 compreende a análise dos resultados, com a filtragem dos dados, mapeamento das sequências contra um transcriptoma de referência, normalização dos dados e teste estatístico para aferir significância na diferença de expressão gênica de um determinado transcrito. A etapa 5 é o estudo biológico dos genes diferencialmente expressos, por meio da anotação gênica, categorização funcional dos genes e análises de vias metabólicas. Adaptado de JAIN (2011), citado por MURATA (2013).

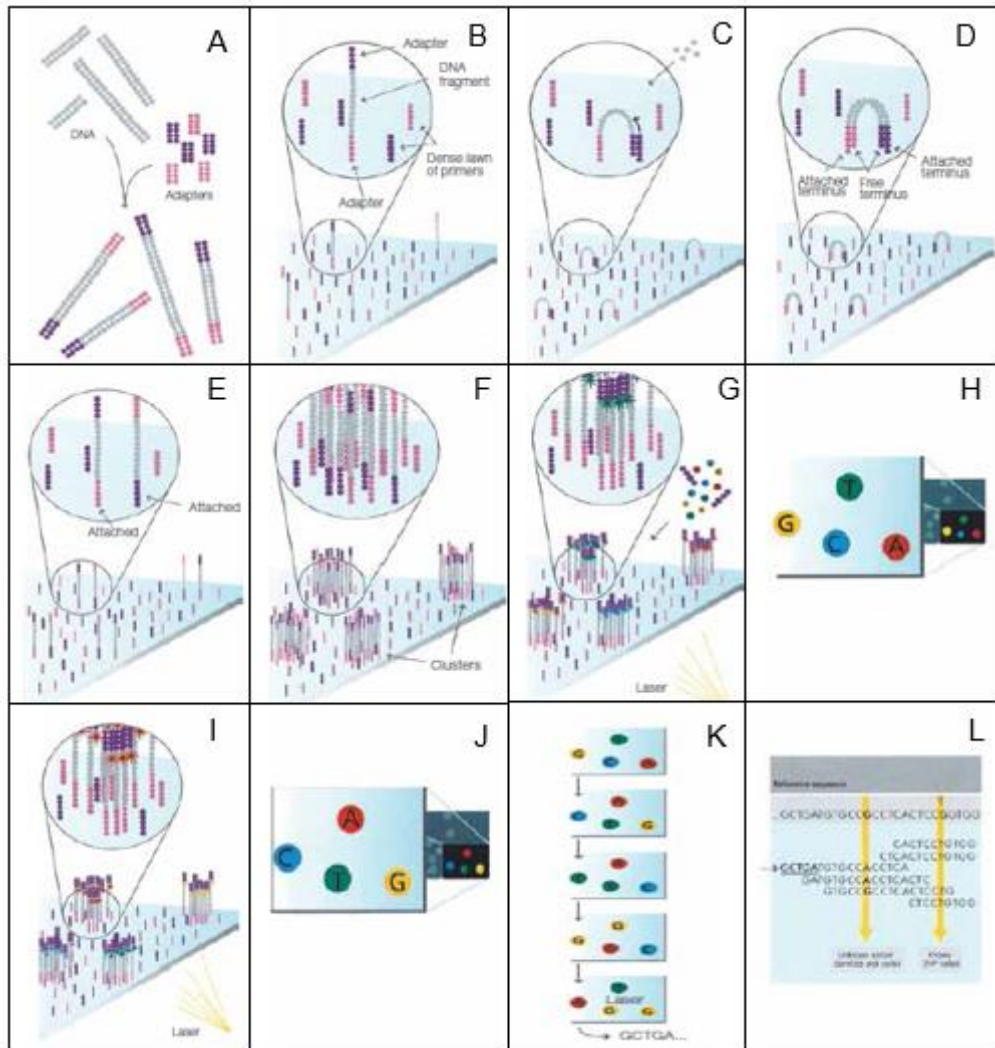
No primeiro ciclo de amplificação, nucleotídeos não marcados são fornecidos para que haja a síntese da segunda fita do fragmento imobilizado no suporte. Após o ciclo de anelamento, o fragmento forma uma estrutura em “ponte” na superfície de sequenciamento e a extensão ocorre, formando a fita complementar também em “ponte” (Figura 5). No ciclo de desnaturação, as fitas são separadas e linearizadas. Esses ciclos são repetidos 35 vezes e assim as cerca de mil cópias geradas de cada

fragmento nessa PCR de fase sólida permanecem próximas umas das outras, formando um *cluster* de sequenciamento (Figura 5).



**Figura 5.** Etapa da clusterização na plataforma Illumina/Solexa. Fonte: Metzker (2010), citado por Cofre (2013).

Etapas de desnaturação são necessárias para a separação dos duplex formados e, nos próximos ciclos de amplificação, nucleotídeos terminadores marcados são fornecidos para as reações de sequenciamento que ocorrem dentro de cada *cluster* (Figura 6). A alta densidade dos *clusters* de sequenciamento possibilita que o sinal de fluorescência gerado com a incorporação de cada um dos nucleotídeos terminadores tenha uma intensidade suficiente para garantir sua detecção exata. Até 50 milhões de *clusters* podem ser produzidos por linha, correspondendo a uma representação satisfatória da biblioteca. Após a incorporação de cada nucleotídeo no fragmento em síntese, a leitura do sinal de fluorescência é realizada. Em seguida, ocorre uma etapa de lavagem para remoção dos reagentes excedentes e remoção do terminal 3' bloqueado e do fluoróforo do nucleotídeo incorporado no ciclo anterior para que a reação de sequenciamento prossiga. A leitura das bases é feita pela análise sequencial das imagens capturadas em cada ciclo de sequenciamento (SHENDURE; JI, 2008).



**Figura 6.** Representação esquemática do princípio tecnológico da plataforma Illumina. O DNA é fragmentado aleatoriamente e ligado a adaptadores A e B em ambas as extremidades (A). As moléculas de DNA fita simples são aderidas por afinidade ao suporte sólido onde estão também aderidos em alta densidade oligonucleotídeos complementares aos adaptadores A e B (B). Durante a etapa de anelamento (C), no primeiro ciclo de amplificação da PCR em fase sólida, o adaptador da extremidade livre da molécula aderida ao suporte encontra seu oligonucleotídeo complementar no suporte, formando uma estrutura em ponte. Uma vez fornecidos os reagentes necessários, a PCR é iniciada utilizando a extremidade 3' livre do oligonucleotídeo como *primer* (C e D). Na etapa de desnaturação (E), a “ponte” é desfeita mediante elevação de temperatura. Repete-se a etapa de anelamento, formando novas estruturas em ponte e iniciando um novo ciclo de amplificação. Após uma série desses ciclos, serão obtidos *clusters* de moléculas idênticas ligadas ao suporte (F). Com a incorporação de nucleotídeos terminadores marcados e excitação a laser (G), é gerado sinal, o qual é captado por dispositivo de leitura e interpretado como um dos quatro possíveis nucleotídeos componentes da cadeia (H). O processo de incorporação de nucleotídeo marcado, excitação e leitura é repetido para

cada nucleotídeo componente da sequência (I, J). A leitura é feita de forma sequencial, o que permite a montagem da sequência completa de cada *cluster* (K). Os dados são alinhados em comparação com uma referência e diferenças nas sequências podem ser identificadas (L). Adaptado de RODRIGUES, C. M., 2011.

Cada corrida de sequenciamento produz “terabytes” de dados em imagem. Os sinais de intensidades dos quatro nucleotídeos são extraídos a partir destas imagens e serão usados para nomear as bases. As imagens são processadas a fim de extrair os sinais numéricos para cada base (ROUGEMONT et al., 2008; MARGUERAT; BÄHLER, 2010). O programa Firecrack (Illumina) de análise de imagens coleta informações para que o programa Bustard converta a informação em uma sequência de DNA de código IUPAC associado com um índice de intensidade e qualidade. Um terceiro programa, chamado Gerald, irá realizar a calibração da qualidade e filtragem dos dados, e finalmente, produzir um arquivo de texto de sequências curtas de leitura em formato FASTQ (WILHELM et al., 2010).

A última etapa é a análise dos dados, que é considerada crítica, porque a enorme quantidade de dados produzidos em cada sequenciamento requer uma infraestrutura adequada para as análises computacionais e locais para o armazenamento dos dados. Além disso, devido ao grande número de sequências, esta etapa é considerada também a mais demorada da tecnologia RNA-Seq (WILHELM et al., 2010).

## **2.6 Montagem *de novo***

Hoje com o constante aprimoramento das técnicas de sequenciamento e a imensa quantidade de dados brutos gerados, ferramentas computacionais para a montagem de genomas e transcriptomas necessitam se tornar cada vez mais robustas e sofisticadas (PASZKIEWICZ; STUDHOLME, 2010; GRABHERR et al., 2011).

Existem basicamente dois caminhos para converter dados de RNA-Seq em sequências de transcritos: com a orientação do mapeamento das sequências obtidas em sequências genômicas já montadas (genoma de referência) ou via montagem *de novo* (HAAS et al., 2013).

A montagem consiste em um conjunto de procedimentos que visa organizar um grande número de sequências curtas em um espaço linear. Geralmente, em projetos de sequenciamento, todo o DNA ou cDNA do organismo de estudo, é fragmentado em milhões de pedaços pequenos, que depois são “lidos” através do sequenciamento automatizado, e posteriormente, para a montagem, todas essas peças são alinhadas umas às outras, tentando identificar as regiões onde os segmentos de leitura se sobrepõem. Quanto mais curtas as sequências, maior a quantidade de sobreposições necessárias para que possa executar esta tarefa (MARTINS, 2013).

O emprego de segmentos de leitura com pareamento de extremidades, isto é, com o sequenciamento das duas extremidades de um mesmo fragmento de DNA (“*paired-end reads*”), separadas por uma distância conhecida, facilita o processo de obtenção de sequências montadas ou contigs (sequências contíguas ou consenso) (MARTINS, 2013).

A montagem *de novo* trabalha em cima da “redundância” das leituras para criar sobreposições e assim montar os transcritos. Nesta abordagem são utilizados os grafos De Bruijn, os quais usam os k-mers (sequências de bases com tamanho k) para representar os nós; os pares de nós são conectados entre si através de uma sobreposição de sequências de exatamente k-1 bases. Alguns montadores realizam montagens com os dados utilizando grafos De Bruijn além de uma etapa de pós-processamento, a fim de fundir os contigs e remover redundância. Esta abordagem, utilizada por programas como Rnnotator, Multiple-k e o Trans-AbySS, permite com que haja perda de isoformas por redundância, enquanto que outros programas, como o Trinity, realizam a reconstrução diretamente a partir dos grafos De Bruijn a fim de montar cada isoforma (ZERBINO; BIRNEY, 2008; MARTIN et al., 2010; ROBERTSON et al., 2010; SURGET-GROBA; MONTOYA-BURGOS, 2010; GRABHERR et al., 2011).

As diferentes estratégias de montagem *de novo* têm vantagens e desvantagens em velocidade e precisão quando comparadas entre si. A montagem *de novo* é quase sempre complexa e difícil, particularmente quando o genoma é grande e o material analisado possui sequências repetitivas (como o caso da cana-de-açúcar), que podem causar falhas na montagem (METZKER, 2010).

Em relação à complexidade do genoma e ao tempo de montagem, montagens *de novo* são mais lentas, consomem muito mais memória de processamento e exigem mais interatividade e atenção do que montagens com referência. Inúmeros são os fatores que afetam a qualidade da montagem, e uma estratégia para a melhoria da qualidade do alinhamento ou montagem tem sido aumentar a cobertura genômica. Além disso, como cada plataforma de sequenciamento produz um padrão diferente de sequências de tamanho e cobertura variável, a união de tipos diferentes de NGS (*Next Generation Sequencing* – Sequenciamento de nova geração) na montagem pode contribuir para corrigir deficiências (MARTINS, 2013).

Aury et al. (2008) relatam uma mistura de sequências obtidas utilizando-se as plataformas Roche 454 e Illumina que resultou em melhora nas montagens *de novo* de genomas microbianos em comparação com as montagens de qualquer uma destas plataformas separadamente.

A otimização dos parâmetros do programa de montagem e a eliminação de sequências de leitura de baixa qualidade também concorrem para a melhoria da montagem de genomas e transcritomas de forma significativa (GARG et al, 2011).

De qualquer forma, essa abordagem, mesmo que trabalhosa, representa uma contribuição significativa para o desenvolvimento de recursos genômicos, estudos funcionais e para apoiar programas de melhoramento de plantas e animais (MARTINS, 2013), além de contribuir para outras pesquisas na área, possibilitando uma maior compreensão e correlação da expressão dos genes e produtos de sua expressão envolvidos nas mais diversas respostas.

## **2.7 Estudos moleculares em cana-de-açúcar na resposta à seca**

O estresse hídrico é um importante fator limitante de crescimento, desenvolvimento e produtividade das plantas cultivadas, levando à modificações morfológicas, bioquímicas, fisiológicas e moleculares nas plantas (JAMAUX, STEINMETZ, BELHASSEN, 1997; BRAY, 1997).

Diversos estudos tem buscado desvendar os mecanismos por trás da tolerância à seca em cana-de-açúcar (SUGIHARTO et al., 2002; ROCHA et al., 2007; CASU et al., 2007; RODRIGUES; DE LAIA; ZINGARETTI, 2009; ISKANDAR et al., 2011; KIDO

et al., 2012; VANTINI, et al., 2015), mas a classificação de genes induzidos sob estresse hídrico e a identificação de genes-chave que contribuem para a tolerância, ainda hoje são um desafio (OLIVEIRA, 2012).

Uma classificação separa os produtos dos genes induzidos pela seca que funcionam na resposta ao estresse, daqueles que atuam no estabelecimento da tolerância. Shinozaki e colaboradores (2003) categorizaram os produtos de genes induzidos pelo estresse em dois grupos: o primeiro inclui proteínas relacionadas à tolerância ao estresse, e o segundo contendo proteínas regulatórias e fatores protéicos envolvidos na transdução de sinal e na expressão de genes responsivos ao estresse.

No primeiro grupo inclui-se as chaperonas, *late embryogenesis abundant* (LEAs) e desidrinas, envolvidas na proteção de estruturas citoplasmáticas durante a desidratação (ROELOFS et al., 2008). No segundo grupo inclui-se proteínas funcionais como as aquaporinas (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 2005), as enzimas chave para biossíntese de osmólitos e açúcares solúveis (CARVALHO, 2008), além de transportadores de açúcares, lipídeos, enzimas de via de desintoxicação e enzimas que atuam no combate ao estresse oxidativo (OLIVEIRA, 2012). Nesse segundo grupo inclui-se também as proteínas regulatórias como os fatores de transcrição, tais como zíperes de leucina, proteínas quinases e fosfatases, enzimas envolvidas no metabolismo de fosfolípidos e moléculas sinalizadoras (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 2005).

Com relação aos estudos moleculares em cana-de-açúcar, formou-se o projeto SUCEST-FUN (<http://sucest-fun.org>), com a finalidade de associar funções aos genes sequenciados pelo SUCEST, projeto genoma da cana. Dentre os 238 mil cDNAs, que compõem o banco de dados SUCEST-FUN, foram identificados genes envolvidos em sinalização, resistência à pragas, resposta a estresses bióticos e abióticos, metabolismo de carboidratos e armazenamento de açúcares, desenvolvimento floral, absorção de nutrientes, regulação do ciclo celular, assimilação de nitrogênio e tolerância a alumínio e metais (VETTORE et al., 2003). Esses ESTs são um excelente recurso para a descoberta de novos genes para o melhoramento da cana.

A partir dos dados do SUCEST e SUCEST-FUN foi iniciado o projeto “Transcriptoma da Cana-de-açúcar”, que produziu microarranjos em lâminas de vidro

contendo 6139 elementos. Nesse projeto foram realizados os primeiros experimentos de análise de expressão gênica em larga escala de plantas de cana, com experimentos de interação com bactérias endofíticas, teor de sacarose, ABA, metil jasmonato, seca, ataque de insetos e resposta a altas concentrações de CO<sub>2</sub> (PAPINITERZI et al., 2005; Rocha et al., 2007; DE SOUZA et al., 2008), que trouxeram grandes informações para a área.

Kido e colaboradores (2012) estudando raízes de cana-de-açúcar tolerantes à seca encontraram, após categorização no Gene Ontology, categorias de genes relacionadas à estímulos hormonais (como etileno, ácido salicílico, ácido jasmônico, ácido abscísico), genes de resposta à escassez de água, estresse oxidativo e osmótico, além de genes ainda não classificados.

Em outro estudo, utilizando a técnica de cDNA-AFLP, Vantini e colaboradores (2015), encontraram 173 fragmentos diferencialmente expressos em resposta ao estresse hídrico em uma cultivar de cana-de-açúcar tolerante à seca. Setenta deles foram clonados, sequenciados, e categorizados. A análise de similaridade utilizando BLAST revelou que 64% dos fragmentos expressos diferencialmente codificavam proteínas não-classificadas (23%), hipotéticas (21%), ou envolvidas em resposta ao estresse (20%), os demais estavam envolvidos na sinalização e transdução de sinal, bioenergética, metabolismo secundário, crescimento e desenvolvimento.

Inúmeros estudos ainda estão sendo desenvolvidos a fim de elucidar os genes responsivos que codificam proteínas estruturais e/ou diretamente relacionadas ao combate do estresse e estabelecimento da tolerância à seca em cana-de-açúcar, o que pode auxiliar na obtenção de genótipos melhorados via manipulação direta (transgenia) ou melhoramento clássico.



### 3 MATERIAL E MÉTODOS

#### 3.1 Material vegetal e condições experimentais

As amostras de tecido vegetal utilizadas foram provenientes do experimento em casa de vegetação vinculado ao projeto multidisciplinar intitulado “Mecanismos fisiológicos em cultivares de cana-de-açúcar submetidas à prolongada limitação hídrica” desenvolvido no laboratório de Química e Bioquímica de Plantas do Departamento de Tecnologia da FCAV/Unesp Jaboticabal – SP, sob supervisão do Prof. Dr. Jairo Osvaldo Cazetta e auxílio da Pesquisadora Dra. Samira Domingues Carlin Cavallari, da Dra. Thais Ramos da Silva e da Mestranda Bruna Robiati Telles.

Todo experimento foi instalado e conduzido em casa de vegetação, em ambiente controlado. O delineamento experimental utilizado foi em blocos casualizados, no esquema fatorial 2×3×3 (duas cultivares (C) × três tensões hídricas do solo (T) × três épocas amostrais), com três repetições, perfazendo um total de 18 unidades experimentais e 54 amostras. Cada unidade experimental foi composta por um vaso de plástico contendo 50 dm<sup>3</sup> de solo.

Foram utilizados dois genótipos comerciais de cana-de-açúcar considerados potenciais em produtividade: SP81-3250 utilizado como padrão de tolerância, e RB855453, suscetível à deficiência hídrica (PINCELLI, 2010).

As mudas de cada um dos genótipos foram obtidas a partir de mini-toletes do colmo de cana-de-açúcar provenientes da Unidade de Pesquisa e Desenvolvimento de Jaú – APTA Pólo Centro Oeste. Estes toletes foram previamente tratados e plantados em recipientes de plástico com capacidade para 0,50 dm<sup>3</sup> contendo areia lavada e peneirada (Figura 7). As mudas originadas da brotação de uma gema foram mantidas em areia, sem qualquer restrição hídrica, durante 30 dias.



**Figura 7.** Mini-toletes de cana-de-açúcar plantados em recipientes de plástico

Após este período, foram selecionadas quanto à sanidade e homogeneidade e transplantadas duas mudas em cada unidade experimental (vaso de plástico com 50 dm<sup>3</sup> de solo), as quais foram cultivadas em solo por 30 dias sem restrição hídrica (período necessário para o estabelecimento e adaptação do sistema radicular), ou seja, até completar 60 dias de idade (Figura 8). A partir desse momento iniciou-se a aplicação dos tratamentos com três distintas tensões hídricas (T1 – sem restrição hídrica, T2 – restrição hídrica moderada, T3 – restrição hídrica severa), as quais foram determinadas por meio dos resultados obtidos com a curva de retenção do solo e apoio de dados da literatura.

O controle do conteúdo de água no solo foi realizado por meio de tensiometria digital a partir do monitoramento diário das leituras das tensões do solo, a fim de se garantir a umidade no nível concernente a cada tratamento. Para isso, foram instalados dois tensiômetros de punção em cada unidade experimental, um na profundidade de 10 cm e outro na de 30 cm, para monitoramento diário da umidade do solo (QUEIROZ, 2010). A estimativa da umidade do solo de cada vaso foi realizada a partir da média das leituras dos tensiômetros instalados nos vasos. A reposição de água foi realizada por meio de um tubo de PVC com perfurações ao longo de toda a sua extensão, instalado no centro geométrico do vaso, garantindo rápida distribuição da umidade do solo em todos os estratos do vaso.



**Figura 8.** Seleção e transplante das mudas para cada uma das unidades experimentais.

Aos 60 dias após o plantio (DAP), houve a diminuição da tensão superficial do solo, dando início à aplicação dos três tratamentos distintos, o controle (sem supressão de rega – 10 a 15 Kpa); estresse moderado (50 a 55 Kpa) e estresse severo (maior que 80 Kpa), os quais foram mantidos e avaliados em três épocas (E): E1 (30 dias após aplicação do estresse), E2 (60 dias após aplicação do estresse) e E3 (90 dias após aplicação do estresse) – Figura 9, constituindo-se das idades de 90, 120 e 150 DAP, período dentro do qual aproximadamente 70 a 80% da produção de cana é definida (SINGH & RAO, 1987).

As avaliações fisiológicas e bioquímicas foram realizadas no laboratório de Química e Bioquímica de Plantas do Departamento de Tecnologia da FCAV/UNESP Jaboticabal – SP, sob supervisão do Prof. Dr. Jairo Osvaldo Cazetta e auxílio da Pesquisadora Dra. Samira Domingues Carlin Cavallari e da aluna de Doutorado Thais Ramos da Silva, aos 30, 60 e 90 dias após o início dos tratamentos (DAT), sendo que nestas mesmas datas foram feitas as coletas de material para a análise da expressão gênica por RNA-Seq.



**Figura 9.** Aplicação das três distintas tensões hídricas em Mpa (T1- sem restrição hídrica, T2- restrição hídrica moderada, T3- restrição hídrica severa). A – Dez dias após aplicação do estresse; B – Sessenta dias após aplicação do estresse; C – Noventa dias após aplicação do estresse.

### 3.2 Coleta das amostras

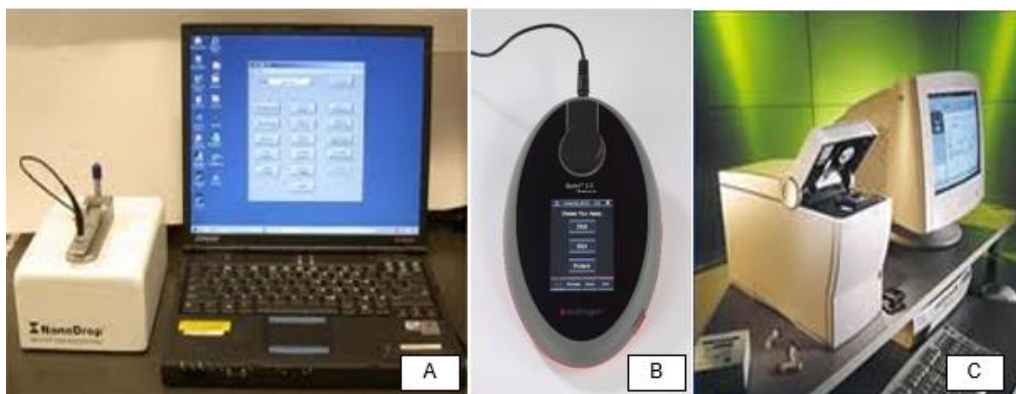
As coletas para avaliações moleculares foram realizadas aos 30, 60 e 90 dias após o início dos tratamentos. Após a retirada da nervura central, fragmentos da folha+1 foram envolvidos em papel alumínio e imediatamente transferidos para nitrogênio líquido. Em seguida, foram armazenados em ultrafreezer (-80°C) até o momento de sua utilização.

### 3.3 Extração de RNA

As amostras foram destinadas à extração de RNA total com uso do kit comercial Purelink RNA Mini Kit (Life Technologies Corporation). A extração de RNA exige cuidados para evitar a contaminação da amostra com RNAses, desta forma, todo material utilizado e todas as soluções usadas para extração foram feitas com água ultra pura tratada com dietilpirocarbonato (DEPC). Foram utilizadas 250 mg de folhas para cada uma das amostras, que foram maceradas em nitrogênio líquido com auxílio de cadinho e pistilo.

As amostras maceradas foram rapidamente transferidas para tubos de microcentrifuga de 1,5 ml contendo tampão de lise fornecido pelo kit e  $\beta$ -mercaptoethanol. As amostras foram homogeneizadas em vortex por 2 minutos e submetidas a uma centrifugação a 3.000xg por 5 minutos a temperatura ambiente. O sobrenadante foi transferido para um novo tubo de microcentrifuga e adicionada a

metade de seu volume de etanol absoluto para a precipitação dos ácidos nucleicos. O tubo foi homogeneizado até o desaparecimento completo do precipitado. A esta solução foi adicionado a coluna Spin Cartridge e os passos seguintes foram realizados segundo as recomendações do fabricante. O RNA purificado foi ressuspendido em 50  $\mu$ l de água ultra pura. A quantificação das amostras de RNA foi determinada no espectrofotômetro NanoDrop ND-1000 (Thermo Fisher Scientific Inc.), através da leitura de absorbância no comprimento de onda 260 nm, 280 nm e 230 nm. As relações 260/280 nm e 260/230 nm estimam a pureza da amostra. A concentração de RNA total em cada amostra foi quantificada no fluorômetro Qubit® 2.0 (Invitrogen) com o kit *Qubit® RNA HS Assay* e sua integridade foi determinada através do Bioanalyzer 2100 (Agilent Technologies), utilizado o Agilent RNA 6000 nano kit, segundo as recomendações do fabricante (Figura 10).



**Figura 10.** Equipamentos utilizados para a quantificação do RNA total. A- NanoDrop ND\_1000; B- Qubit® 2.0 (Invitrogen); C- Bioanalyzer 2100 (Agilent);

### **3.4 Isolamento do RNA mensageiro, montagem das bibliotecas de cDNA e ligação dos adaptadores**

A partir do RNA total das amostras das plantas, o RNA mensageiro (mRNA) foi isolado e purificado pelo kit de preparação de amostras *TruSeq RNA Sample Preparation v2* (Illumina). A quantidade inicial recomendada pela Illumina é de 0,1 a 4  $\mu$ g de RNA total, utilizamos 1,5  $\mu$ g.



O processo de purificação de mRNA baseou-se na ligação da cauda poli-A dessas moléculas ao oligo(dT) ligado à esferas magnéticas. O mRNA purificado foi fragmentado em sequências que variaram de 245 a 300 pb por enzimas presentes no kit. A partir da fragmentação dos RNAs mensageiros, a primeira fita de cDNA foi sintetizada por meio da enzima transcriptase reversa e oligonucleotídeos randômicos. O mRNA molde foi removido para a síntese da segunda fita do cDNA. Para a limpeza das amostras, foi utilizado o kit *Agencourt Ampure beads*. As esferas magnéticas deste kit separaram o cDNA dupla fita do restante dos reagentes.

Os cDNAs com extremidades coesivas resultantes da fragmentação, foram convertidos em extremidades cegas pelo *EndRepair Mix*. Após o reparo das pontas, foi adicionado uma base adenina em cada extremidade 3' dos fragmentos de cDNA para facilitar a ligação dos adaptadores que possuem uma base timina. Os adaptadores responsáveis por identificar as amostras foram ligados às extremidades dos fragmentos de cDNA.

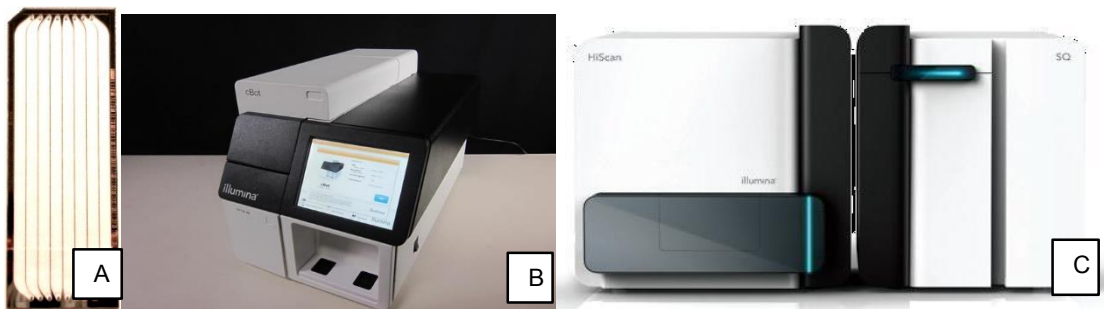
Foi realizada uma reação de polimerase em cadeia (PCR) com oligonucleotídeos específicos para as sequências dos adaptadores para aumentar o número de fragmentos das bibliotecas que continham adaptadores ligados nas duas pontas. Para validação, as bibliotecas foram quantificadas por meio do fluorômetro Qubit® 2.0 e a concentração de cada biblioteca individual foi calculada usando qPCR. Todas as bibliotecas foram normalizadas a 16 pmoles.

### **3.5 Sequenciamento do RNA**

O sequenciamento dos mRNAs via RNA-Seq foi realizado na Facility de sequenciamento da FCAV/UNESP utilizando a plataforma HiScanSQ System Illumina, seguindo as instruções do fabricante para sequenciamento *paired-end* de fragmentos de 50 pb e também na Facility de sequenciamento da ESALQ em Piracicaba utilizando a plataforma HiSeq 2500 Illumina, seguindo as instruções do fabricante para sequenciamento *paired-end* de fragmentos de 100 pb.

### 3.6 Clusterização das bibliotecas de cDNA e sequenciamento

Nesta etapa, as bibliotecas foram inseridas na *flow cell* (lâmina onde acontece a PCR em fase sólida - Figura 11A) para formação de *clusters* e posterior sequenciamento. Para a clusterização foi utilizado o kit *TruSeq SR Cluster Kit v3 – HS* (cBot™ – Figura 11B), seguindo as recomendações do fabricante. O sequenciamento foi realizado na plataforma Illumina (Figura 11C), seguindo as instruções do fabricante para sequenciamento *paired-end* de fragmentos de 50 pb (HiScanSQ) e 100 pb (HiSeq2500).



**Figura 11.** Equipamentos utilizados para clusterização e sequenciamento. A - Flow cell, composta por oito canaletas, com a possibilidade de inserir até 12 amostras em cada uma delas ; B - cBot, equipamento utilizado para clusterização; C -Plataforma HiScanSQ.

### 3.7 Análise dos resultados do sequenciamento

#### 3.7.1 Tratamento das leituras e filtro de qualidade

As leituras geradas no sequenciamento foram agrupadas por bibliotecas em arquivos no formato FASTQ. Essas leituras foram processadas, incluindo a retirada de adaptadores e regiões de baixa qualidade, utilizando o programa SeqyClean (versão 1.8.10) (<https://bitbucket.org/izhbannikov/seqyclean>). Os parâmetros do programa SeqyClean para retirada dos adaptadores foram mantidos, exceto o tamanho mínimo de sequências que foi alterado para 25 nucleotídeos. Para eliminação de regiões de baixa qualidade foi utilizado o índice de qualidade Phred Q30, que refere-se a uma probabilidade de 0,1% de uma base ter sido reconhecida incorretamente.

### 3.7.2 Montagem *de novo* do transcriptoma

As bibliotecas sem adaptadores e somente com as regiões de alta qualidade foram então usadas na montagem do transcriptoma com o programa Trinity (GRABHERR et al., 2013) (versão r20140717).

As montagens foram realizadas separadamente para cada cultivar, ou seja, foi obtido uma montagem da cultivar sensível (RB855453) e uma da cultivar tolerante (SP81-3250), ambas utilizando o programa Trinity. O Trinity inclui o programa Trimmomatic (BOLGER; LOHSE; USADEL, 2014), o qual foi executado para remover bases com qualidade menor que 20 nas extremidades das sequências, e utilizar somente reads maiores que 32 bases. Os parâmetros do Trinity customizados foram a contagem mínima de 3 k-mers para considerá-los válidos e a contagem mínima de reads necessárias para fundir 2 contigs no algoritmo Inchworm do Trinity.

### 3.7.3 Montagem de um transcriptoma referência (Clusterização)

O Trinity contém um padrão de nomenclatura para as sequências de transcritos montados,  $c^{<n^o>}_g^{<n^o>}_i^{<n^o>}$  (por ex.: c1\_g1\_i1), para identificação do componente da montagem (c), gene (g) e isoforma (i). As diferentes isoformas (i) correspondentes ao mesmo gene representam sequências compartilhadas com estruturas distintas, frequentemente promovidas por encadeamentos alternativos de exons (*alternative splicing*). Os diferentes genes, até podem representar partes do mesmo gene, se esse gene não teve cobertura suficiente com o sequenciamento realizado. O componente (c) é uma definição inteiramente operacional do Trinity e representa o agrupamento (cluster) de reads relacionadas, ou seja, que contêm alguma similaridade, e por isso, componentes comuns até podem representar genes relacionados (por ex.: genes parálogos). Sendo assim, o identificador gênico adotado é o componente (c) junto com o gene (g) (por ex.: c1\_g1).

Para montar uma referência de sequências para as duas montagens e com informação gênica comum, ou seja, com identificadores de genes comuns, foi realizado um procedimento de agrupamento (clusterização) com base na similaridade entre as sequências das duas montagens. Essa similaridade entre sequências foi obtida utilizando o programa CD-HIT-EST (LI; GODZIK, 2006), e logo após, as



sequências foram renomeadas de acordo com os grupos formados (clusters) além da informação prévia de genes contida nos identificadores do Trinity.

Desse modo, se um mesmo identificador gênico do Trinity ( $c_{n^o\_g_{n^o}}$ ) aparecer em grupos do CD-HIT distintos, então todos os genes e sequências desses grupos serão reunidos em um novo grupo, o que operacionalmente definimos como gene. Esse gene (grupo de sequências) ganha um novo número identificador ( $c0\_g_{n^o}$ ) onde a informação de componente é ignorada ( $c0$ ).

#### 3.7.4 Mapeamento e cálculo de abundância

As reads foram então mapeadas no transcriptoma referência (Seção 3.7.3) com o programa bowtie2 (v.2.1.0) (LANGMEAD; SALZBERG, 2012), configurado com parâmetros ajustados para aumentar a sensibilidade nas buscas por um alinhamento válido (*very-sensitive*).

A estimativa de abundância dos transcritos montados foi realizada com o utilitário do pacote Trinity (*align\_and\_estimate\_abundance.pl*) utilizando o método implementado no programa eXpress (v.1.5.1). Esses dados de abundância também foram normalizados por FPKM (Fragments Per Kilobase of exon per Million fragments mapped) dentro de cada biblioteca e entre as bibliotecas utilizando TMM (Trimmed Mean of M-values) com *scripts* auxiliares do Trinity.

#### 3.7.5 Identificação dos Genes Diferencialmente Expressos (GDEs)

Para comparar a expressão gênica entre os tratamentos, dois métodos implementados em pacotes R/Bioconductor (v.3.1.2/v.3.0) (GENTLEMAN *et al.*, 2004) foram utilizados: edgeR (v.3.8.5) (ROBINSON *et al.*, 2010; MCCARTHY; CHEN; SMYTH, 2012) e DESeq2 (1.6.3) (LOVE; HUBER; ANDERS, 2014). Ambos contêm implementações de métodos estatísticos baseados em modelos lineares generalizados (MLG) para a detecção da expressão gênica diferencial. Todas as comparações possíveis entre os 18 tratamentos foram realizadas com os dois métodos e selecionadas de acordo com o valor absoluto de  $\log_2$ -fold-change acima de 1 ( $\log_2FC \geq 1$ ) e valor de significância estatística após ajuste de acordo com a taxa

de falsas descobertas (FDR - False Discovery Rate) abaixo de 0,01 (*p-value* ajustado  $\leq 0,01$ ). Além disso, foram excluídos os genes com média de contagem por milhão (CPM) com valores menores que 1 para cada réplica biológica.

### 3.7.6 Anotação dos transcritos

As ORFs (open reading frame – sequência de leitura aberta) e proteínas foram preditas usando o programa TransDecoder (r20140704) (<http://transdecoder.sourceforge.net/>), com a configuração original. As proteínas preditas foram submetidas a uma comparação com o banco de dados UniProtKB/Swiss-Prot (BAIROCH; APWEILER, 1996) utilizando BLAST+ (v.2.2.29) (CAMACHO et al., 2009) com `blastp (-evalue 0.0001 e -max_target_seqs 5)` e anotação de assinaturas de proteínas com o InterProScan (ZDOBNOV; APWEILER, 2001). Para as isoformas que não tiveram ORFs preditas pelo TransDecoder, foi executada uma comparação com `blastx`, contra o UniProtKB/Swiss-Prot.

A fim de complementar a anotação dos transcritos, utilizamos outros bancos de dados de referência, os transcriptomas e proteomas preditos de *Arabidopsis thaliana* e *Sorghum bicolor*, ambos obtidos a partir do repositório da plataforma Phytozome (v.10) (GOODSTEIN et al., 2012), os transcriptomas disponíveis de cana-de-açúcar SoGI (v.3.0) (*Saccharum officinarum* Gene Index) e SCG (SugarCane Genes) (v.1.0) (GRATIVOL et al., 2014), ESTs do banco de dados dbEST (BOGUSKI; LOWE; TOLSTOSHEV, 1993) e mRNAs do GenBank para *Saccharum spp.*, o UniGene de *Saccharum officinarum* e ainda, contra os dados disponíveis de *Miscanthus giganteus* (BARLING et al., 2013). As análises foram realizadas comparando a montagem com os transcriptomas referências no espaço de nucleotídeos (`blastn`) e no espaço de aminoácidos (`blastx/tblastx`), e com os proteomas apenas no espaço de aminoácidos (`blastx/tblastn`). Todas as comparações (parâmetros: `e-value 1e-5` e `-dbsize 500000000`) foram realizadas duas vezes, uma com o transcriptoma montado como item de consulta (*query*) e outra com o banco de dados (*db*). Esse procedimento foi necessário, a fim de se encontrar os melhores correspondentes recíprocos (Reciprocal Best Hits). Para ser reconhecido como um correspondente válido, os alinhamentos com as melhores pontuações (somente os que possuíam pontuação

dentro de um limiar de até 5% abaixo da melhor pontuação), com uma cobertura de alinhamento de no mínimo 50% da sequência consultada e uma identidade maior ou igual a 75% de no mínimo 50 aminoácidos (blastx/tblastx/tblastn) ou 150 nucleotídeos (blastn). As sequências com as variantes transcricionais montadas foram utilizadas nessas comparações, porém a anotação foi realizada por gene.

### 3.7.7 Integração dos dados e análise exploratória

Os dados de abundância, anotação e de expressão diferencial dos genes foram integrados por meio de rotinas automatizadas desenvolvidas em Perl. As análises exploratórias dos dados utilizando agrupamentos hierárquicos foram realizadas com rotinas desenvolvidas em R, com o auxílio de funções de pacotes R/Bioconductor.

### 3.7.8 Análise de agrupamentos de perfis de expressão gênica diferencial

Um agrupamento de perfis de expressão gênica diferencial foi realizada a partir dos resultados de expressão gênica diferencial considerando as comparações entre as cultivares nos 3 tempos. Os valores de  $\log_2$ -fold-change obtidos com edgeR foram selecionados para as 3 comparações (SxT-30, SxT-60 e SxT-90), o método de agrupamento selecionado foi o *K-means*, utilizando um valor de  $k=4$  (número de agrupamentos), definidos por meio da distância euclidiana. Esse agrupamento foi realizado no ambiente R por meio da função `kmeans()`.

## **4 RESULTADOS E DISCUSSÃO**

### **4.1 Material Vegetal**

No projeto multidisciplinar “Mecanismos fisiológicos em cultivares de cana-de-açúcar submetidas à prolongada limitação hídrica”, sob supervisão do Prof. Dr. Jairo Osvaldo Cazetta, foi realizada uma análise fisiológica das cultivares contrastantes utilizadas nesse trabalho (RB855453 e SP81-3250). Foram analisados parâmetros morfológicos como número de perfilhos e de folhas verdes, altura da planta e área foliar (TELLES, 2013), bem como parâmetros fisiológicos como condutância estomática, transpiração, potencial hídrico das folhas, teor de clorofila, assimilação de CO<sub>2</sub>, analisados por Carlin (comunicação pessoal).

A conclusão deste projeto mostrou que a cultivar SP81-3250, tolerante, apresentou maior produtividade e maiores valores na maioria dos parâmetros avaliados. A RB855453, considerada sensível, apresentou valores intermediários de produtividade apesar da taxa fotossintética maior que a cultivar SP em situação de estresse mais prolongado.

As cultivares avaliadas apresentaram respostas morfofisiológicas distintas em relação aos parâmetros analisados evidenciando uma forte influência do genótipo e sua interação com o ambiente. Tais parâmetros mostraram-se eficientes em distinguir as duas cultivares de cana-de-açúcar utilizadas quanto à tolerância e suscetibilidade à seca (TELLES, 2013).

### **4.2 Verificação da concentração e integridade do RNA total extraído**

Para a construção de uma boa biblioteca de RNA - seq foi necessário obter um RNA de boa qualidade. Na Tabela 1 podemos observar os resultados da avaliação das amostras de RNA total, obtidos no Nanodrop. Os resultados do RNA total para a relação A260/A280nm das amostras variou entre 1,97 e 2,22, e a relação A260/230nm variou de 0,33 a 2,38. SAMBROOK; FRITCH; MANIATS (1989) consideram amostras

de boa qualidade quando as leituras de 260/280 apresentam relações de 1,8 a 2,0 e 260/230 quando acima de 2,0. Relações fora destes valores indicam contaminantes na amostra como proteínas ou tiocianato de guanidina, respectivamente. Algumas das amostras analisadas se encontravam dentro desses valores, o que nos permitiu dar continuidade às etapas seguintes do projeto

Uma outra análise das amostras foi realizada com o equipamento Qubit, que permitiu estimar a contaminação por DNA e proteína, além de quantificar o RNA total. Esta técnica utiliza fluoróforos específicos para RNA, DNA e proteína, o que se caracteriza como a principal vantagem deste método, pois possibilita quantificar separadamente os ácidos nucleicos, que acabam sendo “somados” na espectrofotometria. A Tabela 1 mostra as concentrações de RNAs obtidas através do Qubit que variaram de 38,10 a 250,0 ng/μL, resultados que foram suficientes para a construção das bibliotecas e sequenciamento dos transcritos.

**Tabela 1.** Resultados obtidos com a leitura das amostras de RNA total nos três equipamentos: NanoDrop, Qubit e Bioanalyzer

Amostra*	Repetição	NanoDrop		Qubit (ng/μL)	Bioanalyzer RIN
		260/280	260/230		
SSTN3	1	2,10	2,12	162,0	7.9
SSTN3	2	2,16	1,18	200,0	7.8
SSTN3	3	2,07	2,2	43,4	7.9
SSTM3	1	2,12	2,21	200,0	7.9
SSTM3	2	2,10	1,95	140,0	7.8
SSTM3	3	2,08	1,85	150,0	7.8
SSTS3	1	2,09	0,95	56,0	7.7
SSTS3	2	2,12	2,35	120,0	7.7
SSTS3	3	1,97	1,09	81,0	7.4
SSSN3	1	2,12	2,16	63,0	7.9
SSSN3	2	2,10	2,23	170,0	8.0
SSSN3	3	2,11	1,96	162,0	7.4
SSSM3	1	2,08	2,11	71,0	7.9
SSSM3	2	2,10	1,71	91,4	7.8
SSSM3	3	2,06	2,25	140,0	8.0
SSSS3	1	2,15	2,01	87,0	7.7
SSSS3	2	2,05	0,33	38,1	7.3
SSSS3	3	2,09	2,38	250,0	7.3
SSTN6	1	2,14	1,79	156,0	8.0
SSTN6	2	2,04	1,1	176,0	8.0

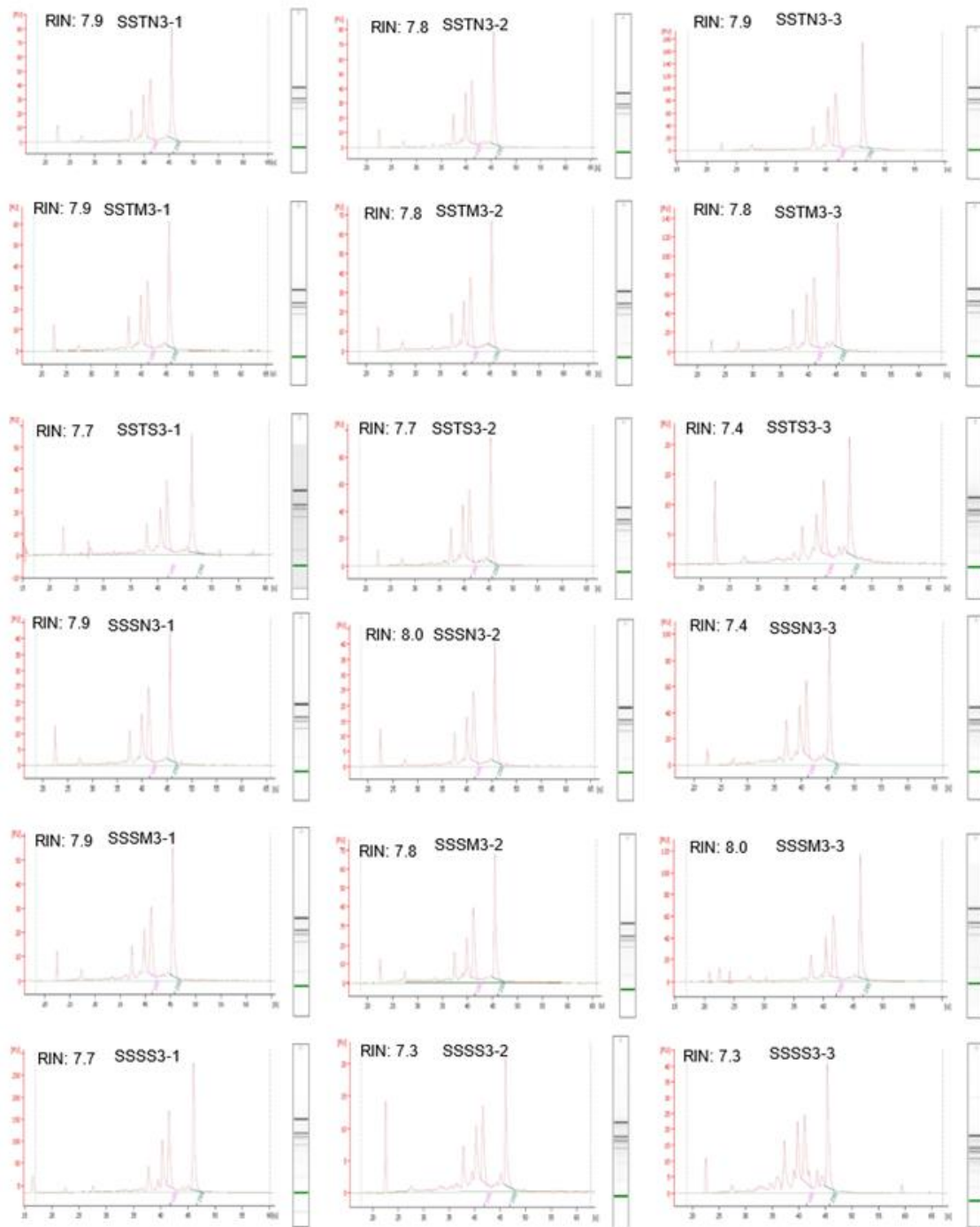
Continuação					
Amostra*	Repetição	NanoDrop		Qubit (ng/μL)	Bionalyzer RIN
		260/280	260/230		
SSTN6	3	2,13	1,36	188,0	7.9
SSTM6	1	2,17	0,73	134,0	7.8
SSTM6	2	2,14	2,25	116,0	8.0
SSTM6	3	2,14	1,7	116,0	8.0
SSTS6	1	2,15	1,94	192,0	7.9
SSTS6	2	2,17	1,57	164,0	7.4
SSTS6	3	2,09	1,37	106,0	7.8
SSSN6	1	2,18	0,97	120,0	8.2
SSSN6	2	2,07	2,21	138,0	8.0
SSSN6	3	2,18	0,46	130,0	7.8
SSSM6	1	2,15	2,29	134,0	7.8
SSSM6	2	2,06	2,22	198,0	8.1
SSSM6	3	2,13	1,53	146,0	8.0
SSSS6	1	2,18	1,27	168,0	8.0
SSSS6	2	2,12	1	77,40	7.6
SSSS6	3	2,13	1,18	192,0	7.9
SSTN9	1	2,16	2,00	97,6	8.0
SSTN9	2	2,07	1,80	134,0	7.8
SSTN9	3	2,18	1,97	124,0	7.8
SSTM9	1	2,20	2,05	114,0	7.8
SSTM9	2	2,15	2,24	81,8	7.8
SSTM9	3	2,18	1,58	106,0	7.8
SSTS9	1	2,22	1,60	128,0	7.8
SSTS9	2	2,17	1,45	164,0	7.9
SSTS9	3	2,09	2,03	86,2	7.5
SSSN9	1	2,17	1,77	140,0	7.8
SSSN9	2	2,18	2,18	160,0	7.8
SSSN9	3	2,13	1,67	136,0	7.8
SSSM9	1	2,18	2,09	156,0	8.2
SSSM9	2	2,19	1,29	54,0	8.0
SSSM9	3	2,20	1,95	192,0	8.0
SSSS9	1	2,08	1,77	118,0	8.1
SSSS9	2	2,18	1,72	132,0	8.1
SSSS9	3	2,16	1,59	134,0	7.4

\*SS- *Saccharum* sp.; S- cultivar sensível (RB855453) / T- cultivar tolerante (SP81-3250); Tratamento: N- não estressado; M- estresse moderado; S- estresse severo; Épocas de estresse: 3 - 30 dias, 6- 60 dias , 9- 90 dias.

Uma outra análise realizada foi avaliar a integridade do RNA total através do equipamento *Agilent 2100 Bioanalyzer (Agilent Technologies)*. A avaliação da integridade nesse equipamento é baseada na razão entre as bandas de RNAs

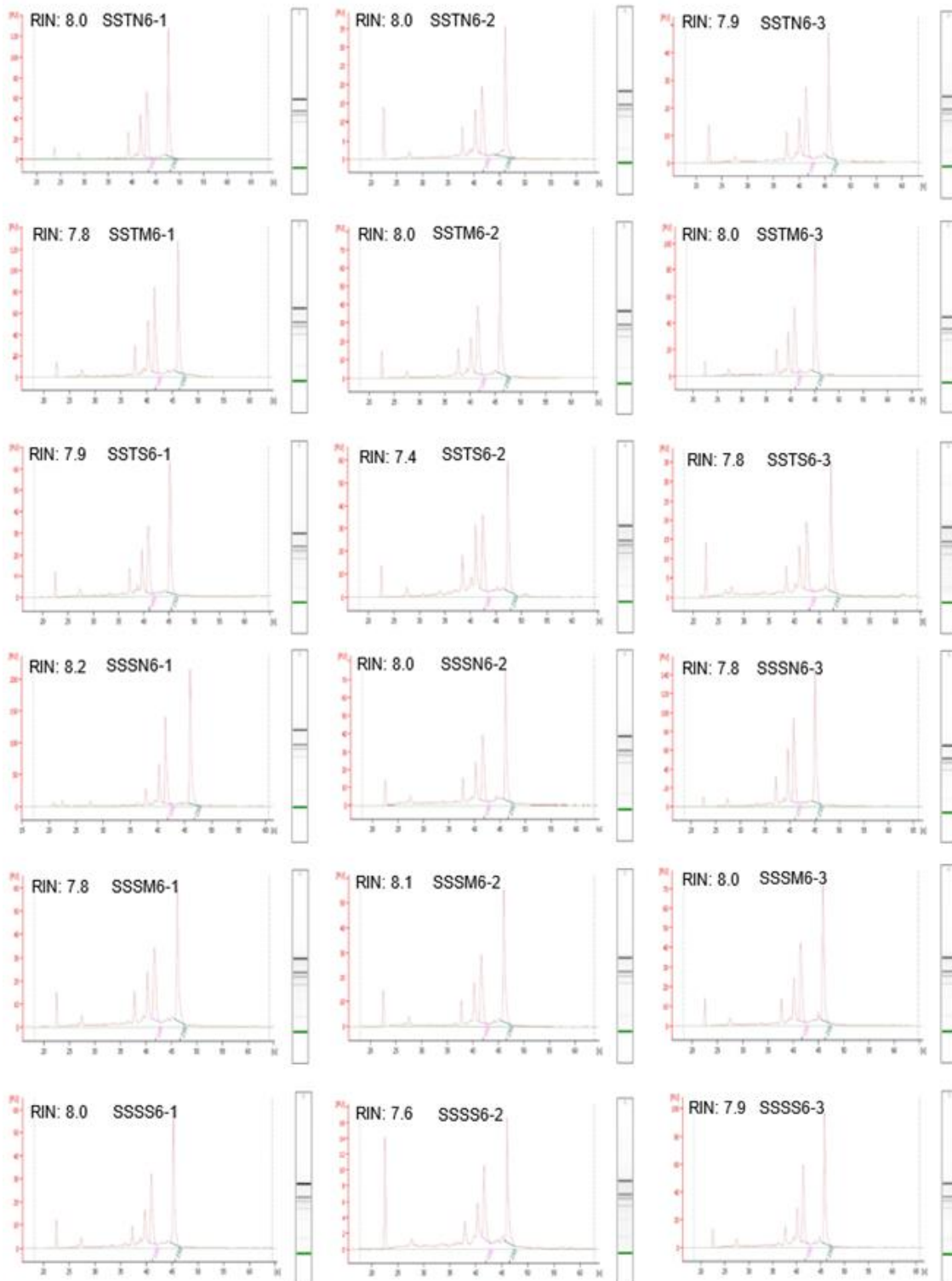
ribossomais 18S e 28S e no sinal de emissão de fluorescência entre os dois picos de detecção das bandas. O software do equipamento expressa a integridade das amostras baseado em um sistema numérico de 1 a 10, onde o 1 representa o perfil mais degradado e o 10 o mais intacto, chamado de *RNA Integrity Number (RIN)*. Portanto, os valores entre 1 e 10 indicam a variação gradual da integridade do RNA total, de mais degradado para mais íntegro (BISELLI,2011).

Os resultados obtidos para os valores de integridade do RNA (RIN) variaram entre 7,3 e 8,2 (Tabela 1 e Figuras 12, 13 e 14), indicando que todas as amostras estavam com uma boa integridade e que poderiam ser utilizadas para o sequenciamento do transcriptoma de cana-de-açúcar por apresentaram a qualidade necessária para análise via RNA-seq. Os eletroferogramas de todas as amostras geradas no equipamento Bioanalyzer, apresentaram dois picos além daqueles baseados na razão entre as bandas de RNA ribossomal 18S e 28S. Uma das possíveis explicações é que esses dois picos menores possam ser de pequenos RNAs ribossomais de cloroplasto, o que já foi descrito na literatura para *Arabidopsis* que apresentava esse mesmo padrão (informações disponíveis no protocolo da *Agilent Technologies*: {HYPERLINK: "http://www.chem.agilent.com/Library/usermanuals/Public/5185-5818\_v4.0.pdf"}).

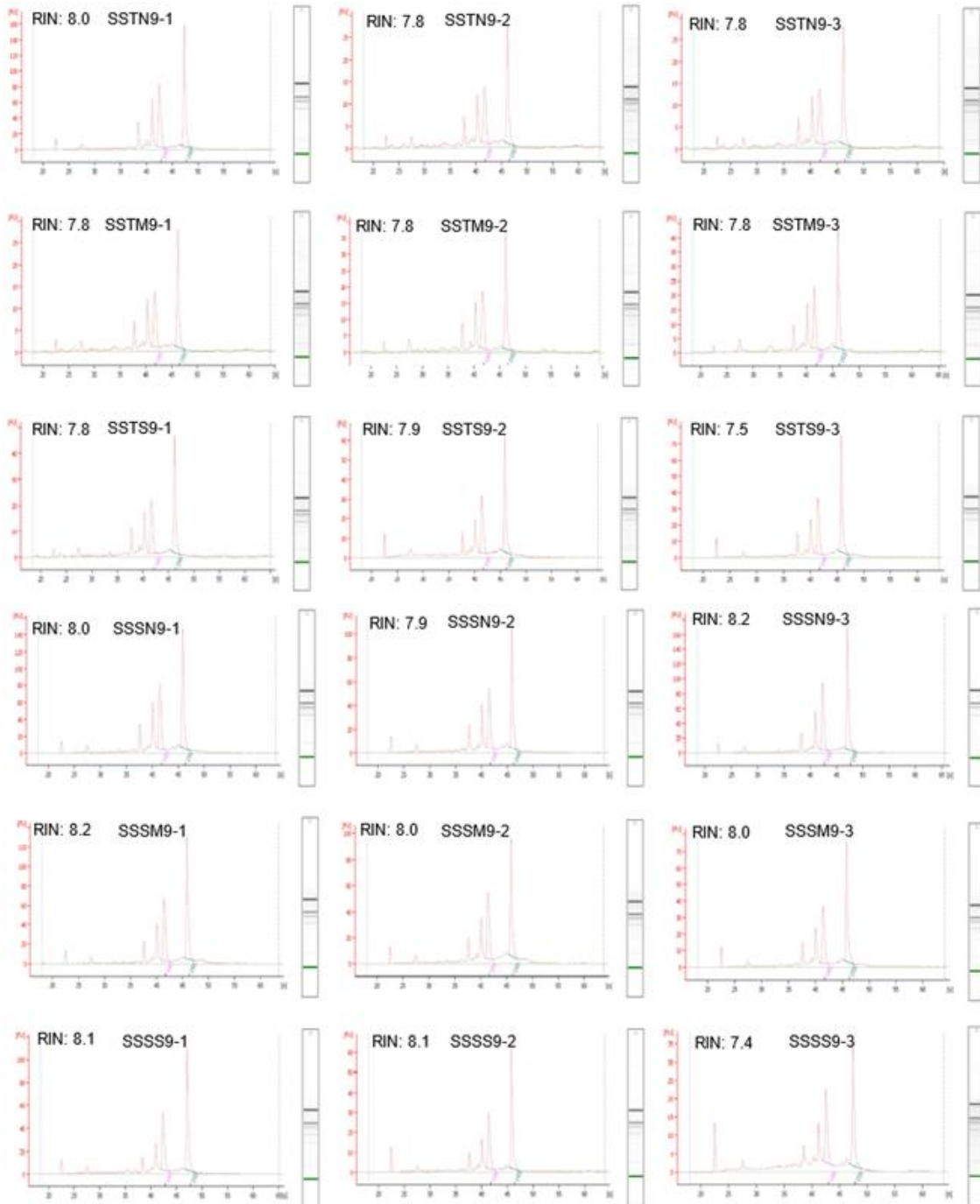


**Figura 12.** Eletroferograma e valor de RIN das amostras de RNA total de folhas de cana-de-açúcar coletadas aos 30 dias de estresse hídrico, gerados pelo Bioanalyzer (*Agilent 2100 Bioanalyzer*).





**Figura 13.** Eletroferograma e valor de RIN das amostras de RNA total de folhas de cana-de-açúcar coletadas aos 60 dias de estresse hídrico, gerados pelo Bioanalyzer (*Agilent 2100 Bioanalyzer*).



**Figura 14.** Eletroferograma e valor de RIN das amostras de RNA total de folhas de cana-de-açúcar coletadas aos 90 dias de estresse hídrico, gerados pelo Bioanalyzer (*Agilent 2100 Bioanalyzer*).

### **4.3 Análise do transcriptoma de folhas de cana-de-açúcar das duas cultivares aos 30, 60 e 90 dias de déficit hídrico**

#### 4.3.1 Tratamento das leituras e filtro de qualidade

Para identificar as diferenças nos mecanismos moleculares de resposta ao estresse hídrico, a tecnologia de RNA-seq foi utilizada, com o sequenciamento em larga escala dos transcritos de RNA (transcriptoma) presentes nas folhas de cana-de-açúcar tolerante (SP81-3250) e sensível (RB855453), nos tempos de 30, 60 e 90 dias após aplicação de déficit hídrico.

Os dados da Tabela 2 mostram o número de leituras geradas para cada uma das amostras, bem como o índice de qualidade e revelam que a quantidade de sequências de baixa qualidade removida foi muito pequena, fazendo com que o número de sequências utilizadas na análise não diminuísse muito quando comparado às sequências brutas (de 1.712.145.471 leituras, o valor passou para um total de 1.631.373.110, entre pareadas e não pareadas), evidenciando o sucesso e boa qualidade do sequenciamento.

**Tabela 2.** Resultado pré e pós-processamento inicial dos dados após o sequenciamento.

<b>Amostra*</b>	<b>Nº de Sequências Pré-Processamento</b>	<b>Qualidade Pré-Proc, (bases Q&gt;=30)</b>	<b>Nº de Sequências Pós-Processamento [PS / UPS]</b>	<b>Qualidade Pós-Proc, (bases Q&gt;=30)</b>
SSTN3-1	18.647.749	87,02%	17.362.876 / 770.016	92,01%
SSTN3-2	24.129.298	85,15%	21.947.827 / 1.153.591	91,05%
SSTN3-3	26.932.633	92.32 %	26.572.302 / 309.541	94,43%
SSTN6-1	16.533.394	56,46%	11.867.917 / 2.148.253	74,21%
SSTN6-2	17.686.387	93,73%	16.758.089 / 472.872	96,62%
SSTN6-3	33.564.272	91.65 %	32.594.671 / 645.201	94,54%
SSTN9-1	49.232.187	85,37%	45.574.668 / 2.206.604	91,02%
SSTN9-2	31.107.404	93.46 %	30.690.302 / 319.665	95,15%
SSTN9-3	58.270.794	84,88%	53.385.711 / 2.923.818	90,97%
SSTM3-1	16.803.821	85,36%	15.541.732 / 764.869	90,84%
SSTM3-2	27.031.164	93.23 %	26.703.462 / 289.237	95,29%
SSTM3-3	47.733.844	69,93%	39.505.095 / 4.240.403	82,05%
SSTM6-1	57.691.801	90,06%	54.722.730 / 1.781.261	93,83%
SSTM6-2	26.594.489	85,65%	24.564.720 / 1.210.347	91,15%
SSTM6-3	14.508.881	92,62%	13.886.313 / 421.227	95,53%
SSTM9-1	16.301.413	93.25 %	15.763.220 / 212.129	99,86%
SSTM9-2	31.756.126	93,11%	30.444.340 / 927.620	95,91%
SSTM9-3	26.325.899	94,06%	25.200.140 / 689.797	96,66%
SSTS3-1	16.816.298	64,11%	13.374.402 / 1.801.375	77,51%
SSTS3-2	19.588.072	94,47%	18.787.571 / 523.371	96,90%
SSTS3-3	19.895.781	86,56%	18.579.249 / 863.560	91,45%
SSTS6-1	9.747.705	94,19%	9.319.799 / 273.445	96,85%
SSTS6-2	30.367.143	68,10%	24.480.225 / 2.945.581	81,54%
SSTS6-3	16.161.803	93.97 %	15.945.083 / 168.495	97,34%
SSTS9-1	25.441.538	86,53%	23.738.953 / 1.052.168	91,58%

<b>Continuação</b>				
<b>Amostra*</b>	<b>Nº de Sequências Pré-Processamento</b>	<b>Qualidade Pré-Proc, (bases Q&gt;=30)</b>	<b>Nº de Sequências Pós-Processamento [PS / UPS]</b>	<b>Qualidade Pós-Proc, (bases Q&gt;=30)</b>
SSTS9-2	60.509.713	58,21%	43.537.533 / 7.346.218	76,51%
SSTS9-3	23.909.879	92,34%	22.681.916 / 718.095	95,49%
SSSN3-1	37.480.176	87,32%	34.918.211 / 1.543.226	92,30%
SSSN3-2	45.350.518	92,65%	43.131.408 / 1.367.349	95,79%
SSSN3-3	25.488.492	90,09%	24.064.391 / 837.671	93,91%
SSSN6-1	29.039.554	85,37%	26.774.945 / 1.476.724	91,07%
SSSN6-2	20.848.855	93,22 %	20.572.037 / 232.620	96,21%
SSSN6-3	96.079.473	54,96%	66.899.978 / 12.309.346	74,34%
SSSN9-1	17.256.307	57,95%	12.686.858 / 2.186.074	74,68%
SSSN9-2	14.342.805	90,44%	13.698.100 / 424.382	93,79%
SSSN9-3	10.580.680	93,03%	10.166.485 / 291.615	95,63%
SSSM3-1	38.911.273	93,26%	38.404.159 / 388.435	95,04%
SSSM3-2	33.933.745	93,06%	32.289.286 / 948.575	95,97%
SSSM3-3	71.347.235	83,60%	65.172.047 / 3.631.983	90,06%
SSSM6-1	52.928.530	86,06%	48.952.197 / 2.374.817	91,56%
SSSM6-2	38.212.092	92,25%	36.421.002 / 1.170.214	95,45%
SSSM6-3	22.917.019	93,06 %	22.433.452 / 361.528	94,77%
SSSM9-1	32.162.909	84,36%	29.590.019 / 1.525.240	90,37%
SSSM9-2	26.905.877	85,80%	25.048.047 / 1.174.794	90,99%
SSSM9-3	8.243.162	85,47%	7.586.580 / 375.340	90,75%
SSSS3-1	20.258.101	86,08%	18.795.417 / 855.861	91,33%
SSSS3-2	17.723.650	92,55%	16.856.170 / 648.912	95,80%
SSSS3-3	62.759.848	59,80%	46.343.765 / 7.598.328	76,89%
SSSS6-1	43.230.312	86,66%	39.823.263 / 1.838.796	92,18%
SSSS6-2	58.256.808	68,74%	47.179.589 / 5.512.213	81,77%
SSSS6-3	32.428.127	93,80%	30.910.668 / 995.007	96,64%
SSSS9-1	37.205.810	93,74 %	36.770.189 / 317.268	96,06%
SSSS9-2	22.522.578	92,46 %	22.156.521 / 262.882	94,31%

Continuação				
Amostra*	Nº de Sequências Pré-Processamento	Qualidade Pré-Proc, (bases Q>=30)	Nº de Sequências Pós-Processamento [PS / UPS]	Qualidade Pós-Proc, (bases Q>=30)
SSSS9-3	32.442.047	92.79 %	32.027.468 / 342.053	94,76%
Total de leituras	1.712.145.471	-	1.631.373.110	-

Nas colunas, em ordem: o nome das amostras, o número de sequências geradas, qualidade Phred (Q>=30) em porcentagem das mesmas, número de sequências após o processamento evidenciando as sequências pareadas (PS) e as sequências não pareadas (UPS) e a qualidade.

\*SS- Saccharum sp.; S- cultivar sensível (RB855453) /T- cultivar tolerante (SP81-3250); Tratamento: N- não estressado; M- estresse moderado; S- estresse severo; Épocas de estresse: 3- 30 dias, 6- 60 dias , 9- 90 dias. Repetição biológica : 1, 2 e 3.

#### 4.3.2 Montagem *de novo* do transcriptoma

A Tabela 3 apresenta a estatística descritiva das duas montagens *de novo* geradas separadamente para cada uma das cultivares. O processo de montagem com o software Trinity resultou para a cultivar tolerante (SP81-3250) em 177.509 transcritos (todas as isoformas gênicas), sendo 93.755 genes distintos com o tamanho médio de 833,06 e um total de 147.875.185 bases. Com relação ao transcrito mais longo, o tamanho médio foi de 715,95 com 67.123.628 bases.

Para a cultivar sensível (RB855453) foram identificados 185.153 transcritos (todas as isoformas gênicas), 101.367 genes distintos com tamanho médio de 816,35 e um total de 151.148.742 bases. O transcrito mais longo apresentou tamanho médio de 700,05 e 70.961.559 bases (Tabela 3).

Tais resultados foram semelhantes aos valores de N50 e porcentagem de GC obtidos na montagem *de novo* de transcriptomas de cana-de-açúcar encontrados no trabalho de Cardoso – Silva (2014), utilizando o mesmo programa para a montagem.

**Tabela 3.** Estatística descritiva da montagem dos transcriptomas das cultivares tolerante e sensível de cana-de-açúcar utilizando o programa Trinity.

	<b>Total de genes</b>	<b>Total de transcritos</b>	<b>Porcentagem de GC</b>	<b>N50*</b>	<b>Mediana</b>	<b>Média</b>	<b>Número total de bases</b>
<b>Cultivar Tolerante (SP81-3250)</b>							
<b>Genes</b>	93.755	177.509	42,86	1.262	539	833,06	147.875.185
<b>Transcrito mais longo</b>	-	-	-	1.126	417	715,95	67.123.628
<b>Cultivar Sensível (RB855453)</b>							
<b>Genes</b>	101.367	185.153	42,71	1.233	524	816,35	151.148.742
<b>Transcrito mais longo</b>	-	-	-	1.072	412	700,05	70.961.559

\*N50: Número em que 50% dos genes / isoformas estão acima desse valor e 50% abaixo desse valor.

#### 4.3.3 Identificação dos Genes Diferencialmente Expressos (GDEs)

A Tabela 4 representa os genes diferencialmente expressos identificados após a montagem *de novo*, comparando os tratamentos (cultivares e tempos de coleta após restrição hídrica) aos pares. Podemos observar que, quando comparados os tratamentos da mesma cultivar o número de genes diferencialmente expressos foi menor do que quando as cultivares contrastantes foram comparadas. Isso confirma nossas hipóteses de que a maior variação a nível gênico acontece entre cultivares contrastantes.

Para complementar os resultados e comparar os perfis de expressão de cada uma das 54 amostras foi construído um *HeatMap* (Figura 15), onde foi possível visualizar o comportamento das cultivares frente ao estresse a que foram submetidas e agrupá-las de acordo com a semelhança.

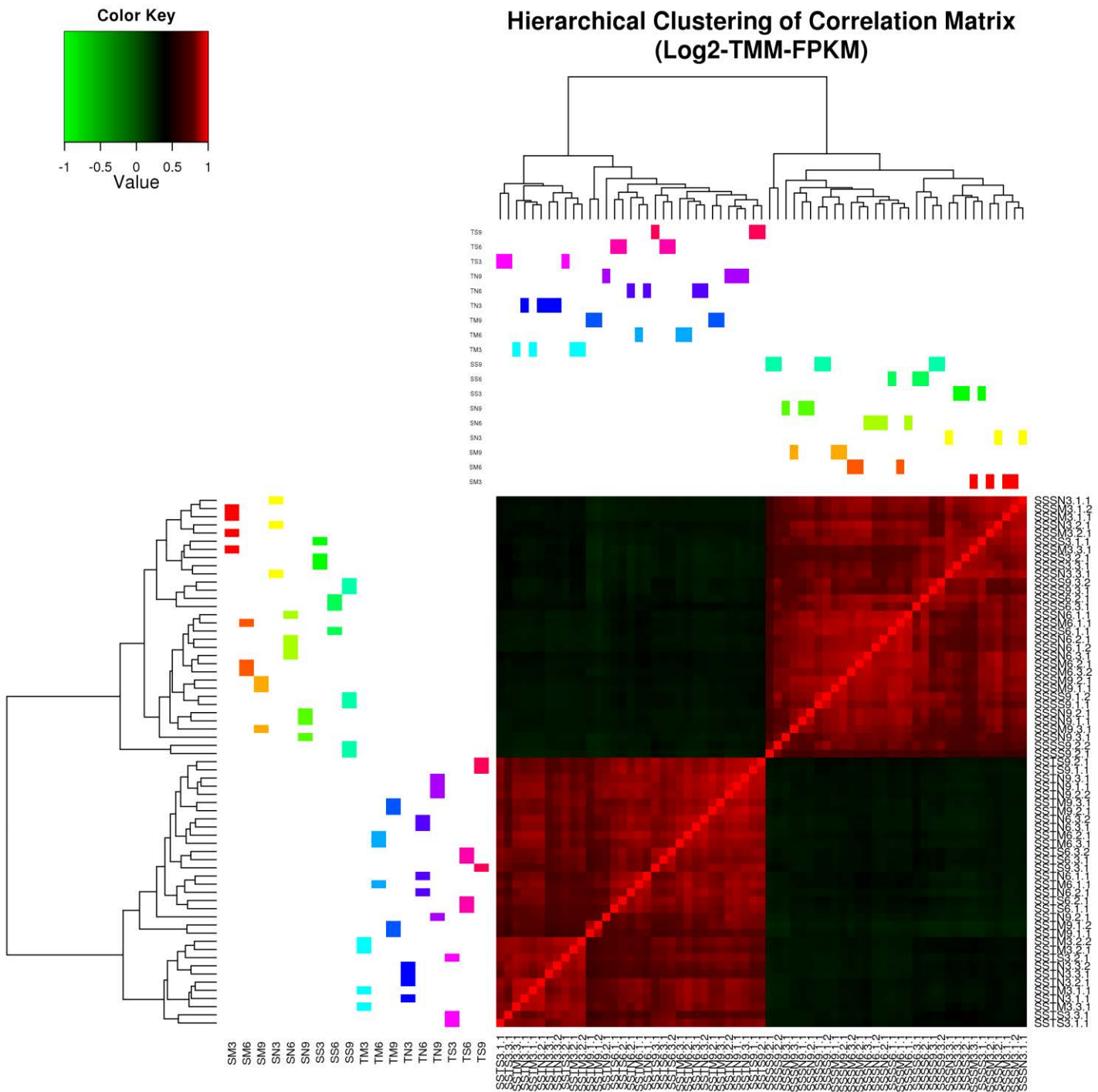
A Figura 15, assim como a Tabela 4, mostram que houve uma maior correlação entre os tratamentos da mesma cultivar. Quanto maior for a similaridade da expressão na comparação (par-a-par), maior será o valor, ou seja, mais próximo de 1, e a cor característica será o vermelho. Por outro lado, quanto menor a similaridade da expressão (valores próximos de zero ou -1) a cor característica será o verde. Dessa forma, é possível concluir que a maior similaridade acontece dentro da mesma cultivar, enquanto as maiores diferenças aparecem quando comparadas as cultivares contrastantes.



**Tabela 4.** Número de genes diferencialmente expressos entre os tratamentos (linha x coluna). Os valores representam os genes diferencialmente expressos induzidos no tratamento representado na linha quando comparado ao da coluna. Valores acima da mediana estão em tons de verde e valores abaixo da mediana em tons de vermelho, valores próximos à mediana (3.028,5) estão em tons de amarelo.

Tratamento*	SSSN3	SSSN6	SSSN9	SSSM3	SSSM6	SSSM9	SSSS3	SSSS6	SSSS9	SSTN3	SSTN6	SSTN9	SSTM3	SSTM6	SSTM9	SSTS3	SSTS6	SSTS9
SSSN3	0	2453	662	0	1001	848	0	1039	3266	3710	7060	6459	3742	6002	6835	3358	6888	6699
SSSN6	496	0	12	949	0	29	1405	10	588	5849	3714	4156	5732	3093	4729	5142	3820	3610
SSSN9	78	8	0	268	2	1	433	34	286	4289	3905	3222	4284	3166	3563	3823	3852	3138
SSSM3	0	4095	1286	0	1745	1482	7	1606	4196	4250	<b>8578</b>	7586	4421	7334	<b>8445</b>	3509	<b>8244</b>	7416
SSSM6	210	0	15	432	0	15	789	2	309	4967	3374	3768	4939	2949	4123	4431	3498	3488
SSSM9	131	77	2	410	11	0	1045	28	94	5323	3808	3141	5032	3163	3227	4559	3880	3046
SSSS3	0	3261	1176	3	1444	1760	0	759	2459	4123	<b>8066</b>	6901	3953	6613	<b>7819</b>	3208	6843	6779
SSSS6	408	66	180	612	3	52	453	0	93	5269	4002	4085	5286	3460	4770	4331	3336	3621
SSSS9	820	838	251	1073	263	99	823	58	0	6046	5309	4439	5582	4342	4752	4588	4194	3229
SSTN3	3074	7066	5276	3591	5307	5670	3501	5033	7461	0	2127	1403	1	1353	2355	2	2331	2288
SSTN6	4299	3835	4095	5567	3196	3625	6626	3410	5982	1286	0	26	1283	0	139	1124	78	126
SSTN9	4230	4628	3759	5291	3788	3455	5823	3658	5277	1049	67	0	1009	55	20	729	348	40
SSTM3	3194	6318	4999	3693	4929	5234	3299	4560	6529	2	1724	1042	0	814	1501	0	1615	1592
SSTM6	3969	3390	3568	5024	3027	3378	5339	3118	5027	940	0	47	730	0	121	612	38	52
SSTM9	4447	5329	4046	5790	4215	3551	6537	4148	5800	1593	352	28	1190	164	0	1171	647	62
SSTS3	3030	7239	5316	3375	5469	5567	2898	4799	6780	12	2650	1411	6	1252	2650	0	1701	1308
SSTS6	4347	4003	4094	5345	3349	3705	5069	2907	4559	1537	48	169	1256	21	471	675	0	4
SSTS9	4591	4336	3586	5406	3693	3374	5592	3219	4044	1857	218	44	1331	86	72	657	25	0

\*SS- *Saccharum* sp.; S- cultivar sensível (RB855453) /T- cultivar tolerante (SP81-3250); Tratamento: N- não estressado; M- estresse moderado; S- estresse severo; Épocas de estresse: 3 - 30 dias, 6- 60 dias , 9- 90 dias.



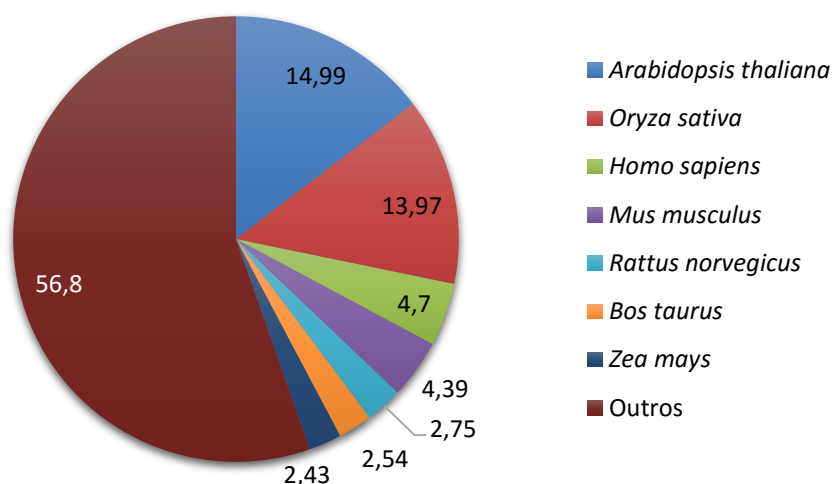
**Figura 15.** Agrupamento hierárquico (método de distância baseado na correlação de Pearson/método de ligação completa) e HeatMap comparando os valores de expressão ( $\log_2$ -FPKM) normalizados (TMM) entre as 54 amostras para os genes identificados como diferencialmente expressos em ao menos uma comparação.

Analisando-se os cinco maiores valores de genes diferencialmente expressos da Tabela 4, que são 8578, 8445, 8244, 8066 e 7819 , é possível perceber que não existe um tempo de estresse que sobressai, mas os cinco maiores valores apareceram aos 30 dias na cultivar sensível quando comparado à tolerante nas diferentes épocas.

#### 4.3.4 Anotação dos transcritos

Os transcritos foram anotados contra o banco de dados UniProtKB/Swiss-Prot utilizando BLAST+ (blastp/blastx). Como pode ser observado no gráfico (Figura 16), a maior porcentagem de anotação se deu contra os organismos *Arabidopsis thaliana* e *Oryza sativa*, com 14,99 % e 13,97% de isoformas anotadas, respectivamente, sendo que *Oryza sativa* pertence à família das gramíneas (Poaceae), mesma família da cana-de-açúcar.

Por outro lado, quando confrontamos os dados da montagem com o SwissProt, houve uma porcentagem de alinhamento com outros organismos, tais como *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* e outros organismos e plantas que não pertencem à família *Poaceae*. Estabelecemos três hipóteses sobre este fato: i) O banco de dados utilizado possui uma maior frequência de genes para *Homo sapiens* (20.198), *Mus musculus* (16.711), *A. thaliana* (13.888) e *Rattus norvegicus* (7.921), por exemplo, enquanto a cana-de-açúcar possui uma frequência de 304 genes depositados (UNIPROTKB, 2015); ii) A segunda possibilidade diz respeito aos genes ortólogos, que são genes que derivam de um ancestral comum através de um evento de especiação, e na maioria dos casos, possuem funções equivalentes em diferentes organismos. Isso explicaria a anotação da cana-de-açúcar com os demais organismos (GOMES, 2013); iii) Por fim, existe também a possibilidade de contaminação com tais organismos nas nossas amostras.



**Figura 16.** Distribuição (%) das espécies mais representadas dentre as proteínas correspondentes aos genes da montagem do transcriptoma de cana-de-açúcar anotados com o banco de dados UniProtKB/Swiss-Prot.

Para complementar a anotação dos transcritos, utilizamos outros bancos de dados de referência da cana-de-açúcar e organismos próximos (Tabela 5). Pudemos observar que a maior porcentagem de anotação se deu contra variedades de cana-de-açúcar utilizadas no trabalho de Cardoso-Silva (2014) com 40% de genes anotados, seguido pelo banco de ESTs de cana-de-açúcar com 21,07%. As anotações contra *M. giganteus*, *S. bicolor* e demais bancos relacionados à cana apresentaram cerca de 20% de genes anotados.

A menor porcentagem de anotação apareceu contra *A. thaliana* e sequências de RNA mensageiro, com 0,24% e 0,41% de genes anotados, respectivamente.

**Tabela 5.** Resultados da anotação do transcriptoma da cana-de-açúcar contra os diferentes bancos de dados.

Banco de dados de referência	Número de genes anotados na montagem	Número de genes anotados na montagem (%)	Número de registros únicos anotados na montagem	Número de registros únicos anotados na montagem (%)	Total de registros do banco de dados de referência
PRINTS	3.236	2,43	480	0,36	2.106
Pfam	21.214	15,96	3.858	2,90	14.831
InterPro	21.195	15,94	4.609	3,47	27.002
ProDom	202	0,15	101	0,08	1.894
SMART	6.355	4,78	564	0,42	1.008
TIGRFAM	2.077	1,56	664	0,50	4.284
ProSiteProfiles	8.943	6,73	605	0,46	1.087
KEGG	1.573	1,18	108	0,08	-
Reactome	1.209	0,91	52	0,04	-
MetaCyc	1.014	0,76	457	0,34	-
UniPathway	1.002	0,75	138	0,10	-
SwissProt	5.213	3,92	15.224	11,45	547.357
<b>Transcriptomas</b>					
<b>BlastN/BlastN</b>					
<i>S. bicolor</i>	25.577	19,24	19.497	14,67	33.032
<i>A. thaliana</i>	320	0,24	351	0,26	27.416
<i>M. giganteus</i>	22.762	17,12	24.076	18,11	61.055
CARDOSO-SILVA et. al, 2014	53.678	40,38	51.697	38,89	72.269
SOGI	27.245	20,50	41.374	31,13	121.342
SCG	9.002	6,77	11.061	8,32	36.871
<i>S. officinarum</i> (UNIGENE)	15.841	11,92	12.398	9,33	15.394
<i>Saccharum</i> (ESTs)	28.007	21,07	105.915	79,68	285.138
<i>Saccharum</i> (mRNA)	541	0,41	555	0,42	744

<b>Continuação</b>					
<b>Banco de dados de referência</b>	<b>Número de genes anotados na montagem</b>	<b>Número de genes anotados na montagem (%)</b>	<b>Número de registros únicos anotados na montagem</b>	<b>Número de registros únicos anotados na montagem (%)</b>	<b>Total de registros do banco de dados de referência</b>
<b>tBlastX/tBlastX</b>					
<i>S. bicolor</i>	18.800	14,14	14.838	11,16	33.032
<i>A. thaliana</i>	3.449	2,59	3.772	2,84	27.416
<i>M. giganteus</i>	19.118	14,38	19.867	14,94	61.055
CARDOSO-SILVA et. al, 2014	53.684	40,38	48.971	36,84	72.269
SOGI	25.783	19,40	38.734	29,14	121.342
SCG	7.641	5,75	9.330	7,02	36.871
<i>S. officinarum</i> (UNIGENE)	15.667	11,79	12.111	9,11	15.394
<i>Saccharum</i> (ESTs)	26.538	19,96	101.068	76,03	285.138
<i>Saccharum</i> (mRNA)	600	0,45	535	0,40	744
<b>Proteomas</b>					
<b>BlastX/tBlastN</b>					
<i>S. bicolor</i>	16701	12,56	15606	11,74	33.032
<i>A. thaliana</i>	3512	2,64	4255	3,20	27.416

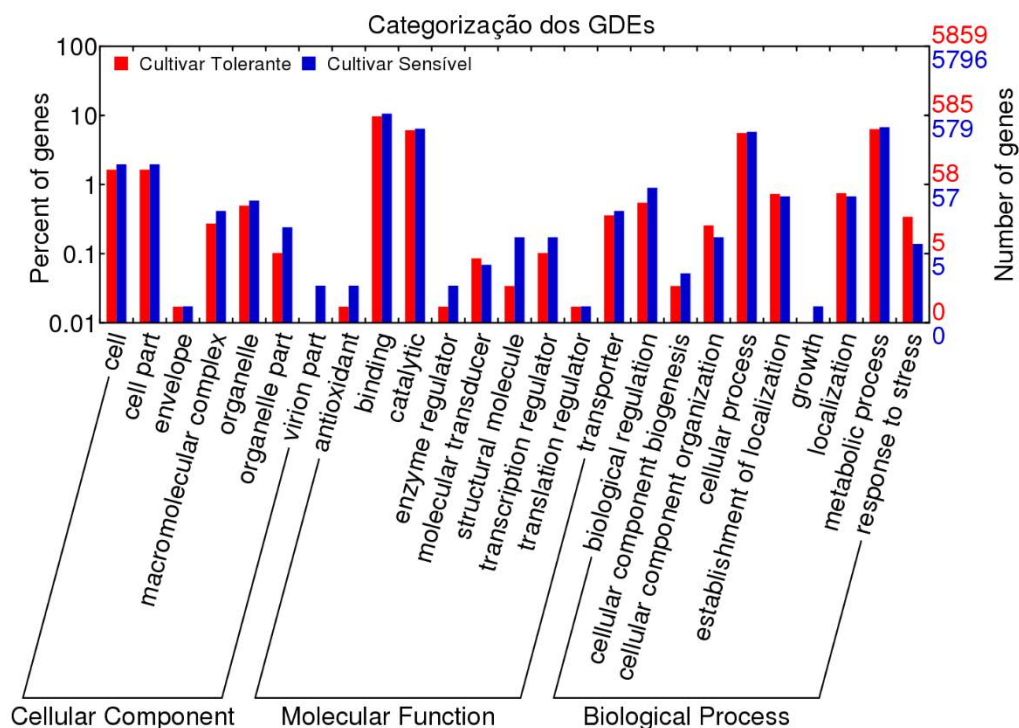
#### 4.3.5 Padrão geral da expressão diferencial das cultivares

Após o levantamento do número de genes diferencialmente expressos (GDEs), foi realizada uma categorização destes genes comparando de maneira geral as cultivares utilizando-se o resultado da anotação do InterProScan e o programa WEGO (YE et al., 2006). Nessa comparação inicial não levou-se em consideração nenhum tempo e nível de estresse; a comparação foi feita entre as cultivares.

O programa WEGO permite criar um gráfico através de uma classificação fornecida pelo Gene Ontology (GO), facilitando assim a visualização da anotação (YE et al., 2006). O Gene Ontology classifica a anotação em três categorias distintas: Processos biológicos, Função molecular e Componente celular (ASHBURNER et al., 2000).

Como pode ser verificado no gráfico da Figura 17, foram identificados 11.655 genes diferencialmente expressos, sendo 5.796 genes diferencialmente expressos induzidos na cultivar sensível, e 5.859 genes diferencialmente expressos induzidos na cultivar tolerante.

Além disso, é possível observar que algumas classes se destacaram por apresentarem um maior número de GDEs, mas sem diferença significativa entre as cultivares. As classes funcionais mais representadas (5 a 10% de genes diferencialmente expressos induzidos em tolerante e sensível) foram ligação e atividade catalítica, na categoria função molecular, e processo celular e processos metabólicos na categoria processos biológicos.



**Figura 17.** Categorização dos genes diferencialmente expressos comparando as duas cultivares, tolerante e sensível, submetidas ao déficit hídrico prolongado. Gráfico gerado pelo Programa WEGO, através da classificação fornecida pelo Gene Ontology.

#### 4.3.5.1 Genes diferencialmente expressos induzidos na cultivar tolerante (SP81-3250)

Dentre os GDEs induzidos identificados (Tabela 6), a cultivar tolerante apresentou um GDE induzido relacionado à transdução molecular, a *histidina quinase – domínio de transferência de fosfato (Hpt)*. A maioria dos sistemas de transdução de sinal de procaríotos e eucariotos usa esquemas de fosfotransferência envolvendo dois componentes conservados, uma proteína histidina quinase (HK) e uma proteína reguladora de resposta (RR). A HK é regulada por estímulos ambientais, e realiza a autofosforilação de um resíduo de histidina, criando um grupo fosfato de alta energia que é posteriormente transferido para um resíduo de aspartato no domínio RR. A fosforilação induz uma alteração conformacional que resulta na ativação de um



domínio associado, efetuando respostas que podem causar alterações na fisiologia celular, muitas vezes regulando a expressão do gene (STOCK; ROBINSON; GOUDREAU, 2000).

Além destes dois domínios de sinalização, alguns sistemas de dois componentes contêm um outro domínio de sinalização comum, denominado domínio de transferência de fosfato (Hpt) que contêm histidina. A grande maioria dos domínios Hpt de organismos procariotos são parte do domínio de multi-quinases, enquanto que em eucariotos são encontrados em proteínas separadas e independentes. Domínios Hpt possuem aproximadamente 120 aminoácidos de comprimento e contêm um resíduo de histidina, capaz de participar em reações de transferência de fosfato. Eles possuem finalidade dupla, atuando como receptores e doadores de fosfato. A presença de domínios Hpt e a utilização de sistemas de várias etapas sugerem outros pontos de regulação da via de sinalização (STOCK; ROBINSON; GOUDREAU, 2000).

Na Tabela 6 observamos na cultivar tolerante, a indução de um gene do complexo de proteínas Exo70 que participa do processo de exocitose. A exocitose é um processo no qual as membranas de vesículas intracelulares fundem-se com a membrana plasmática levando a uma continuidade entre o meio intra e extracelular. Assim, realiza duas funções: a fusão das membranas vesicular e plasmática, e a liberação dos conteúdos intravesiculares para fora da célula. Desta forma, desempenha função essencial para o crescimento e desenvolvimento da planta, principalmente para o crescimento apical em plantas (MORAES, 2005).

Para esse processo de exocitose é necessário o complexo *exocyst*. Esse complexo é formado por oito subunidades: SEC3, SEC5, SEC6, SEC8, SEC10, SEC15, EXO70, e EXO84, que variam em tamanho e são recrutados por diferentes mecanismos (LI et al., 2010).

Acredita-se que a subunidade SEC3 é independente do citoesqueleto de actina, enquanto a EXO70 é parcialmente dependente da actina F (actina filamentosa) (FINGER et al., 1998; BOYD et al., 2004). A Actina F aciona muitos processos celulares, incluindo a morfogênese, divisão e migração celular (HOLMES, 2009).

Em nosso trabalho, a expressão de genes relacionados a Exo70 pode estar relacionada com a maior migração e movimentação de células, responsáveis por

desencadear respostas à deficiência hídrica como o alongamento da raiz, proporcionando maiores condições de sobrevivência diante do estresse.

A cultivar tolerante apresentou um gene relacionado à uma família de DNA helicase, a Pif1 (Tabela 6). As DNA helicases desempenham papéis essenciais na replicação, reparação e recombinação de DNA em todos os organismos (KNOLL; PUCHTA, 2011).

A maioria das helicases de vários organismos têm sido classificadas em duas superfamílias, SF1 e SF2, com base na conservação de sete motivos de helicase. As helicases Pif1 são membros da superfamília de SF1 e têm sido encontradas em todos os eucariotos, de levedura a humano, partilhando um elevado grau de conservação (BOULÉ; ZAKIAN, 2006), sendo essenciais para a manutenção do genoma nuclear e mitocondrial (BOCHMAN; SABOURIA; ZAKIANA, 2010).

No trabalho realizado por Umate, Tuteja e Tuteja (2010) foram identificados em *Arabidopsis* e arroz 115 e 113 genes, respectivamente, pertencentes à família helicase. Dos 113 genes encontrados em arroz, 64 foram homólogos à *Arabidopsis*, leveduras e humanos. Os resultados revelaram que a expressão de genes helicase é profundamente regulada sob várias condições de estresse, fornecendo uma base para a caracterização mais profunda de cada tipo helicase.

A indução de genes da família helicase na cultivar tolerante em cana-de-açúcar sugere uma relação com o estresse hídrico. Esse resultado está de acordo com o que foi descrito no trabalho de Knoll e Puchta (2011), que demonstrou que os genes da família helicase em plantas parecem exercer um papel predominante na modulação das respostas ambientais.

**Tabela 6.** Categorização dos genes diferencialmente expressos induzidos na cultivar tolerante (SP81-3250), em comparação à cultivar sensível (RB855453)

<b>Categoria</b>	<b>Classe</b>	<b>Genes</b>	<b>Anotação InterPro</b>	<b>logFC (DESeq)</b>	<b>logFC (EdgeR)</b>
<b>Função Molecular</b>	<b>Transdução Molecular</b>	c0_g125596	Signal transduction histidine kinase, phosphotransfer (Hpt) domain	-1,1692	-1,1474
		c0_g85180	7TM GPCR, serpentine receptor class g (Srg)	-1,5643	-1,5218
		c0_g36381	Ionotropic glutamate receptor	-1,9778	-1,9567
		c0_g57464	PAS fold;Serine-threonine/tyrosine-protein kinase catalytic domain;PAS domain;Protein kinase domain;Serine/threonine/dual specificity protein kinase, catalytic domain	-1,5823	-1,5418
		c0_g130103	Siganl transduction histidine kinase EnvZ-like, dimerisation/ phosphoacceptor domain	-7,7462	-7,0977
<b>Processos Biológicos</b>	<b>Resposta a estresse</b>	c0_g21585	UspA;Exocyst complex protein Exo70	-1,6920	-1,6675
		c0_g23007	UspA;Exocyst complex protein Exo70	-2,3371	-1,1692
		c0_g44106	DNA helicase Pif1 like;Helitron helicase-like domain;DNA helicase	-1,6174	-1,5816
		c0_g47588	UspA;Exocyst complex protein Exo70	-1,1647	-1,1282
		c0_g91470	DNA polymerase beta, thumb domain;DNA polymerase beta, palm domain;DNA polymerase family X, beta-like;DNA polymerase lambda, fingers domain;DNA polymerase	-1,2157	-1,1789
		c0_g105210	Endonuclease V	-1,0041	-1,0084
		c0_g72187	U-box domain-containing protein kinase family protein;U-box domain-containing protein kinase family protein	-1,1862	-1,1393
		c0_g10016	DNA helicase Pif1 like	-1,5998	-1,5842
		c0_g106663	DNA helicase;DNA helicase Pif1 like	-5,1519	-4,9713
		c0_g23843	Mitochondrial carrier protein;Helitron helicase-like domain;DNA helicase;DNA helicase Pif1 like;Mitochondrial substrate/solute carrier	-2,7447	-2,7273
		c0_g4556	DNA helicase;DNA helicase Pif1 like	-4,3910	-4,2934
		c0_g58503	DNA photolyase, FAD-binding/Cryptochrome, C-terminal;DNA photolyase, N-terminal	-1,3012	-1,2611
		c0_g59009	Helitron helicase-like domain;DNA helicase;DNA helicase Pif1 like	-1,7492	-1,7054
		c0_g68641	DNA helicase;DNA helicase Pif1 like	-8,6507	-7,9816
		c0_g78701	DNA helicase;DNA helicase Pif1 like	-2,4557	-2,4179
		c0_g94988	DNA helicase Pif1 like	-2,5039	-2,4778
c0_g97444	DNA recombination and repair protein Rad51, C-terminal;DNA recombination and repair protein RecA, monomer-monomer interface;DNA recombination and repair protein	-2,1700	-2,1262		
c0_g434	Bet v I domain	-6,5840	-6,5369		

<b>Continuação</b>					
<b>Categoria</b>	<b>Classe</b>	<b>Genes</b>	<b>Anotação InterPro</b>	<b>logFC (DESeq)</b>	<b>logFC (EdgeR)</b>
<b>Processos Biológicos</b>	<b>Organização de componentes celulares</b>	c0_g65939	Membrane insertase OXA1/ALB3/YidC	-1,1405	-1,1064
		c0_g56360	Histone H3-K9 methyltransferase, plant; Pre-SET domain; Pre-SET zinc-binding sub-group; SRA-YDG; Post-SET domain; SET domain	-2,8221	-2,7335
		c0_g61487	Pre-SET domain;Pre-SET zinc-binding sub-group;SRA-YDG	-3,0358	-3,0262
		c0_g10016	DNA helicase Pif1 like	-1,5998	-1,5842
		c0_g106663	DNA helicase;DNA helicase Pif1 like	-5,1519	-4,9713
		c0_g23843	Mitochondrial carrier protein;Helitron helicase-like domain;DNA helicase;DNA helicase Pif1 like;Mitochondrial substrate/solute carrier	-2,7447	-2,7273
		c0_g44106	Helitron helicase-like domain;DNA helicase;DNA helicase Pif1 like	-1,6174	-1,5816
		c0_g4556	DNA helicase;DNA helicase Pif1 like	-4,3910	-4,2934
		c0_g47588	DNA helicase Pif1 like	-1,1647	-1,1282
		c0_g59009	Helitron helicase-like domain;DNA helicase;DNA helicase Pif1 like	-1,7492	-1,7054
		c0_g68641	DNA helicase;DNA helicase Pif1 like	-8,6507	-7,9815
		c0_g78701	DNA helicase;DNA helicase Pif1 like	-2,4557	-2,4179
		c0_g94988	DNA helicase Pif1 like	-2,5039	-2,4778
		c0_g62432	Microtubule-associated protein, MAP65/Ase1/PRC1	-1,1763	-1,1281

#### 4.3.5.2 Genes diferencialmente expressos induzidos na cultivar sensível (RB855453)

Após a categorização dos GDEs (Figura 17) observamos que a cultivar sensível obteve expressão de genes induzidos em duas categorias em que a cultivar tolerante não expressou nenhum gene diferencialmente expresso, como fragmentos de vírus e crescimento. A Tabela 7 apresenta tais categorias com os genes relacionados com estas funções.

No presente trabalho, a cultivar sensível expressou genes induzidos relacionados a proteínas ribossomais, que catalisam a ligação entre dois aminoácidos na síntese de proteínas.

Essas proteínas que compõem o rRNA são nomeadas de acordo com a subunidade do ribossomo a que pertencem. A subunidade pequena, é nomeada do S1 ao S31, e a subunidade grande é nomeada no L1 ao L44 (RAMAKRISHNAN; MOORE, 2001; MAGUIRE; ZIMMERMANN, 2001).

A função dessas proteínas é organizar e estabilizar a estrutura terciária do rRNA, e exercem atividades cruciais de decodificação e transferência de peptídeos baseados no RNA, desempenhando papel ativo nas funções que podem ter evoluído para agilizar o processo de síntese protéica (SANYAL; LILJAS, 2000; RAMAKRISHNAN; MOORE, 2001).

A cultivar sensível apresentou proteínas induzidas localizadas na subunidade pequena do ribossomo (S7, S13, S17 e S23) e também proteínas localizadas na subunidade maior (L1, L6 e L24). Essas proteínas podem estar desempenhando papel fundamental de resposta, aumentando a velocidade da síntese de proteínas, como forma de contornar o estresse após sua percepção.

A cultivar sensível apresentou ainda genes induzidos relacionados ao crescimento como o *COBRA*, *plant* (Tabela 7). As proteínas da família COBRA estão envolvidas na determinação da orientação da expansão celular, com papel importante na deposição de celulose na parede celular (REA; VATAMANIUK; RIGDEN, 2004).

A parede celular primária de plantas combina propriedades únicas de força e rigidez além de plasticidade e viscosidade, que auxiliam no crescimento e proteção das células. Tais propriedades derivam da natureza dos diferentes componentes da parede: microfibrilas de celulose, embebida em uma matriz amorfa e hidratada (65%

de água) de hemicelulose, pectinas e glicoproteínas. A expansão celular depende diretamente do controle local destas estruturas (ROUDIER et al., 2005).

As microfibrilas de celulose representam o componente mais importante no controle da expansão celular, e sua ausência resulta numa perda rápida de crescimento (ROUDIER et al., 2005).

Expressando as proteínas da família COBRA, provavelmente, a cultivar sensível pode evitar a ausência de celulose na parede celular, como forma de manter o crescimento mesmo sob déficit hídrico.

Na cultivar sensível, houve expressão induzida também do gene GrpE (GroP-like gene E), que codifica uma proteína que auxilia na troca de nucleotídeos das proteínas de choque térmico – HSP70 (Heat Shock Proteins) (Tabela 7).

A resposta ao estresse celular é um mecanismo evolutivamente conservado, que permite às células responderem às condições metabólicas e ambientais adversas. Essa resposta é representada, em nível molecular, pela indução da síntese de diferentes tipos de proteínas com função de proteção celular, quando submetidas a estresses bióticos e/ou abióticos, as quais estão associadas à expressão de proteínas de choque térmico (ARAP, 2003).

As chaperonas moleculares interagem com diversos substratos protéicos para auxiliar na adequada conformação tridimensional de proteínas recém-formadas. Elas exercem papel importante durante o estresse celular prevenindo a formação de proteínas com conformação tridimensional incorreta (BORGES; PEROTO; RAMOS, 2001; VANTINI et al., 2015).

Conseqüentemente as proteínas de choque térmico auxiliam na recuperação celular após o estresse reparando proteínas mal-formadas ou então promovendo sua degradação. As proteínas reguladas pela glicose (glucose-regulated proteins – GRP) são exemplos de chaperonas conhecidas (ARAP, 2003).

Genes codificantes de GRPs são regulados por auxina, ácido abscísico, seca, frio, ferimentos, luz, temperatura, salinidade, encharcamento e patógenos, com padrão de expressão bem diversos, variando em função do desenvolvimento e do tecido/órgão analisados (POERSH, 2011).

Bocca et al (2005) afirmam que os genes codificantes de GRPs são finamente regulados durante o desenvolvimento bem como sob muitas influências externas, tendo em muitos casos regulação tecido-específica.

A atividade da co-chaperona GrpE é atuar como fator de liberação de nucleotídeos do ADP da Hsp70 podendo trabalhar como um sensor térmico. (HARRISON, 2003; CAGLIARI, 2009).

Portanto, no presente trabalho, a expressão induzida do gene *GrpE* na cultivar sensível sugere que a planta esteja fazendo uso também deste mecanismo de proteção celular, quando submetida ao déficit hídrico.

Neste trabalho, a cultivar sensível apresentou também um gene induzido relacionado à atividade antioxidante, com anotação no InterPro de *Plant peroxidase; Haem peroxidase* (Tabela 7). Em plantas submetidas a condições de seca ocorre um aumento na produção de espécies reativas de oxigênio (EROS), como o radical superóxido ( $O_2^-$ ), peróxido de hidrogênio ( $H_2O_2$ ), radical hidroxila ( $OH^-$ ) e o oxigênio singlet ( $O_2^1$ ) (JALEEL et al., 2007), que causam peroxidação dos lipídeos e danos nos pigmentos, proteínas e ácidos nucleicos (PRICE; HENDRY, 1991; SMIRNOFF, 1993).

As peroxidases exercem importantes funções na defesa antioxidante removendo peróxidos, incluindo o peróxido de hidrogênio. Ao remover os peróxidos, as plantas podem reduzir a formação de radicais livres e minimizar os danos causados pelo estresse oxidativo, através de enzimas de varredura de radicais livres como a catalase (CAT), superóxido dismutase (SOD), guaiacol peroxidase (POD) e enzimas do ciclo ascorbato-glutationa, como a ascorbato peroxidase (APx) (PURVIS; SHEWFELT, 1993).

As peroxidases estão relacionadas com os processos de crescimento e diferenciação celular, e mudanças morfogênicas em resposta aos estresses físico, químico e biológico. O aumento da atividade desta enzima em plantas submetidas a estas condições pode ser fator determinante da capacidade de adaptação dessas plantas (PIZA; LIMA; BRASIL, 2003).

**Tabela 7.** Categorização de genes mais expressos na cultivar sensível (RB855453), em comparação à cultivar tolerante (SP81-3250)

<b>Categoria</b>	<b>Classe</b>	<b>Genes</b>	<b>Anotação InterPro</b>	<b>logFC (DESeq)</b>	<b>logFC (EdgeR)</b>	
<b>Componente Celular</b>	<b>Fragmentos de vírus</b>	c0_g63988	Potato leaf roll virus readthrough protein	8,2803	8,1956	
		c0_g79161	RNA-directed RNA polymerase, luteovirus;Luteovirus group 1 coat protein;Peptidase S39B, luteovirus;RNA-directed RNA polymerase, catalytic domain;Luteovirus Orf2 putative replicase 1;Luteovirus VPG protein;Sugarcane yellow leaf virus (SCYLV), Orf0	8,0120	8,0114	
		c0_g30020	Coatomer beta subunit, C-terminal; Coatomer beta subunit, appendage platform domain	1,5992	1,6383	
<b>Função Molecular</b>	<b>Fragmentos de organelas</b>	c0_g14836	Mitochondrial brown fat uncoupling protein; Mitochondrial substrate/solute carrier	1,4909	1,5550	
		c0_g123018	UDP-galactose transporter; Nucleotide-sugar transporter	3,0680	3,0845	
		c0_g128466	Endonuclease/exonuclease/phosphatase;Reverse transcriptase domain; Reverse transcriptase zinc-binding domain; E2F/DP family, winged-helix DNA-binding domain	1,6759	1,7143	
		co_g101910	Mitochondrial glycoprotein	1,2996	1,3343	
		c0_g103803	Gamma-tubulin complex component protein	1,7088	1,7620	
		c0_g95023	Gamma-tubulin complex component protein	2,4154	2,4554	
		c0_g10853	Mitochondrial substrate/solute carrier	5,5451	5,3218	
		c0_g78369	Ribossomal protein L1/ribosomal biogenesis protein; Ribossomal protein L1, bacterial-type Zinc Finger,	1,7986	1,8616	
		c0_g16433	Ribosomal protein S7, bacterial/organellar-type;Ribosomal protein S7 domain	1,3320	1,3969	
		c0_g71959	RING-type; Replication termination factor 2, RING-finger; Photosystem II PsbP, oxygen evolving complex	1,2463	1,2979	
	c0_g118783	Rad21/Rec8-like protein, C-terminal, eukaryotic; Rad21/Rec8-like protein, N-terminal	1,0760	1,1271		
	<b>Regulação de enzimas</b>	c0_g74851	GrpE nucleotide exchange factor	1,4263	1,4823	
		c0_g90447	PRONE domain	1,1674	1,1878	
	<b>Molécula Estrutural</b>		c0_g30020	Coatomer beta subunit, C-terminal;Coatomer beta subunit, appendage platform domain	1,5992	1,6383
			c0_g12654	Ribosomal protein S13	1,0825	1,1339
			c0_g16433	Ribosomal protein S7 domain;Ribosomal protein S7, bacterial/organellar-type	1,3320	1,3969



<b>Continuação</b>						
<b>Categoria</b>	<b>Classe</b>	<b>Genes</b>	<b>Anotação InterPro</b>	<b>logFC (DESeq)</b>	<b>logFC (EdgeR)</b>	
<b>Função Molecular</b>	<b>Molécula Estrutural</b>	c0_g49460	Ribosomal protein L24		1,3599	1,3934
		c0_g74560	Ribosomal protein L6, bacterial-type;Ribosomal protein L6, alpha-beta domain		1,5295	1,5996
		c0_g76629	Ribosomal protein S17, archaeal/eukaryotic;Ribosomal protein S17		1,9920	2,0431
		c0_g78369	Ribosomal protein L1, bacterial-type;Ribosomal protein L1/ribosomal biogenesis protein		1,7986	1,8616
		c0_g86191	Ribosomal protein S23, eukaryotic/archaeal;Ribosomal protein S12/S23		2,0223	2,0870
	<b>Antioxidante</b>	c0_g21368	Plant peroxidase;Haem peroxidase, plant/fungal/bacterial		1,7599	1,8178
		c0_g4744	EF-hand domain;NADPH oxidase Respiratory burst		1,0145	1,0411
	<b>Regulação da transcrição</b>	c0_g10585	DNA-binding WRKY		1,6786	1,8537
		c0_g123089	AP2/ERF domain		1,0240	1,0548
		c0_g128466	Endonuclease/exonuclease/phosphatase;Reverse transcriptase domain;Reverse transcriptase zinc-binding domain;E2F/DP family, winged-helix DNA-binding domain		1,6759	1,7143
		c0_g132647	DNA-binding WRKY;HAT dimerisation domain, C-terminal;Zinc finger, BED-type;hAT-like transposase, RNase-H fold		6,3247	6,2331
		c0_g16589	Basic-leucine zipper domain		1,5280	1,5664
		c0_g38455	DNA-binding WRKY		1,8150	1,8181
		c0_g66075	Leucine-rich repeat;DNA-binding WRKY;NB-ARC		4,0763	4,1019
		c0_g68180	DNA-binding WRKY		1,5651	1,6073
	c0_g93997	Ankyrin repeat-containing domain;Transcription factor, K-box;Zinc finger, RING-type;Ankyrin repeat;Transcription factor, MADS-box		1,3047	1,3390	
	<b>Processos Biológicos</b>	<b>Biogênese de componentes celulares</b>	c0_g84427	COBRA, plant		7,7824
c0_g83750			Plus-3 domain, subgroup;Plus-3;SWIB/MDM2 domain		3,1022	3,1510
	<b>Crescimento</b>	c0_g84427	COBRA, plant		7,7824	7,8048
	<b>Regulação biológica</b>	c0_g11417	Reverse transcriptase zinc-binding domain;Zinc finger, CCHC-type;Endonuclease/exonuclease/phosphatase;Reverse transcriptase domain;Diacylglycerol kinase, catalytic domain		3,4322	3,4802
		c0_g12346	Protein kinase C-like, phorbol ester/diacylglycerol-binding domain;Diacylglycerol kinase, accessory domain;Diacylglycerol kinase, catalytic domain		1,5163	1,5687
		c0_g121330	ferritin 4		1,1236	1,1578

<b>Continuação</b>					
<b>Categoria</b>	<b>Classe</b>	<b>Genes</b>	<b>Anotação InterPro</b>	<b>logFC (DESeq)</b>	<b>logFC (EdgeR)</b>
<b>Processos Biológicos</b>	<b>Regulação biológica</b>	c0_g119598	Thioredoxin domain;Protein kinase domain;Serine/threonine/dual specificity protein kinase, catalytic domain	8,3272	7,9082
		c0_g13282	Thioredoxin domain;Thioredoxin-like fold	1,0866	1,1160
		c0_g39167	Thioredoxin domain;Thioredoxin-like fold	1,8731	1,9081
		c0_g69132	Thioredoxin domain;Thioredoxin-like fold	7,3280	7,1310
		c0_g77585	Thioredoxin domain;Thioredoxin-like fold	1,8140	1,8791
		c0_g80930	Thioredoxin domain	8,7958	8,8462
		c0_g93185	Thioredoxin domain;Tetratricopeptide repeat;Tetratricopeptide repeat-containing domain	1,4393	1,4758
		c0_g129677	DNA helicase;DNA helicase Pif1 like	8,2697	7,9021
		c0_g6174	DNA helicase;Replication factor A, C-terminal;DNA helicase Pif1 like	1,5994	1,6381
		c0_g133695	Small GTP-binding protein domain;Small GTPase superfamily, Ras type;Ran GTPase;Small GTPase superfamily, Rho type;Small GTPase superfamily;Small GTPase superfamily, Rab type;Small GTPase superfamily, ARF type	1,0526	1,1135
		c0_g26345	NAF domain;NAF/FISL domain;Protein kinase domain;Serine/threonine/dual specificity protein kinase, catalytic domain	1,0467	1,0990
		c0_g63037	Small GTP-binding protein domain;Small GTPase superfamily, Ras type;Ran GTPase;Small GTPase superfamily, Rho type;Small GTPase superfamily;Small GTPase superfamily,	1,8076	1,8621
		c0_g92967	Small GTP-binding protein domain;Small GTPase superfamily, Ras type;Small GTPase superfamily, Rho type;Small GTPase superfamily;Small GTPase superfamily, Rab type	1,7374	1,7706
		c0_g9926	Signal transduction histidine kinase, core;Signal transduction histidine kinase EnvZ-like, dimerisation/phosphoacceptor domain;Histidine kinase-like ATPase, C-terminal	2,0598	2,1005
		c0_g10896	Ribosome-inactivating protein	1,0865	1,1078
		c0_g11010	Ribosome-inactivating protein	1,2970	1,3205
		c0_g1993	Ribosome-inactivating protein	7,7932	7,6943
		c0_g32030	Ribosome-inactivating protein;Protein kinase domain;Serine/threonine/dual specificity protein kinase, catalytic domain	3,8288	3,8894
		c0_g43491	Ribosome-inactivating protein	2,4544	2,5098
		c0_g105521	Auxin response factor	1,1259	1,1361
		c0_g108500	NAC domain;Zinc finger, BED-type	8,5004	8,0737
		c0_g112736	Zinc finger, C2HC5-type	1,7371	1,7331
		c0_g123573	NAC domain;Zinc finger, BED-type	7,1341	6,6617
		c0_g125381	B3 DNA binding domain;Auxin response factor	1,0201	1,0342
c0_g130760	NAC domain	1,3994	1,4303		

<b>Continuação</b>					
<b>Categoria</b>	<b>Classe</b>	<b>Genes</b>	<b>Anotação InterPro</b>	<b>logFC (DESeq)</b>	<b>logFC (EdgeR)</b>
<b>Processos Biológicos</b>	<b>Regulação biológica</b>	c0_g16367	Mitochondrial transcription termination factor	1,1112	1,1579
		c0_g26214	NAC domain	1,9261	1,9545
		c0_g32050	Homeobox domain;ELK;KNOX2;Homeobox KN domain	1,0625	1,0701
		c0_g35974	NAC domain	1,4074	1,4815
		c0_g48604	AUX/IAA protein;Aux/IAA-ARF-dimerisation	1,0581	1,0589
		c0_g15188	Zinc finger-XS domain;XS domain	1,9657	1,9825
		c0_g38286	Uncharacterised domain XH;XS domain	1,2280	1,2596

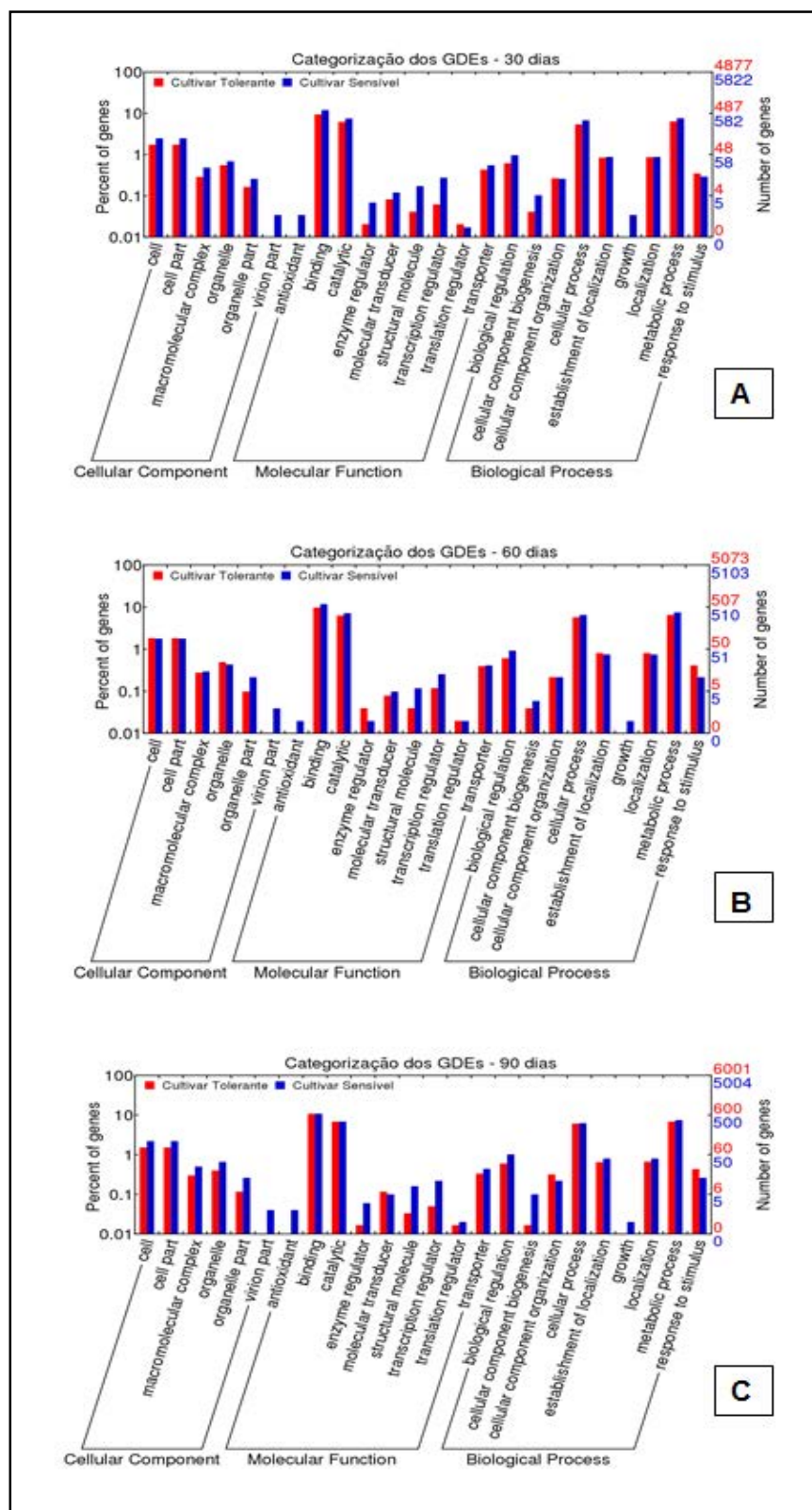
#### 4.3.6 Expressão gênica diferencial aos 30, 60 e 90 dias de estresse

Na Figura 18, podemos observar o resultado da categorização dos 15.966 genes diferencialmente expressos (GDEs induzidos) ao longo do tempo (30, 60 e 90 dias de déficit hídrico) nas duas cultivares.

Ao observar a categorização em cada um dos três tempos estudados, verificamos aos 30 dias um maior número de genes induzidos na cultivar sensível, sendo que essa mesma cultivar apresenta GDEs em categorias que a cultivar tolerante não possui GDEs induzidos. Essas categorias são crescimento, antioxidante e virulência, as quais se mantêm apenas na cultivar sensível nas três épocas, com pequenas alterações na quantidade de genes envolvidos em tais categorias. Aos 30 dias, a cultivar sensível apresenta uma quantidade maior ou igual de GDEs induzidos em comparação com a cultivar tolerante na maioria das categorias (Figura 18 A)

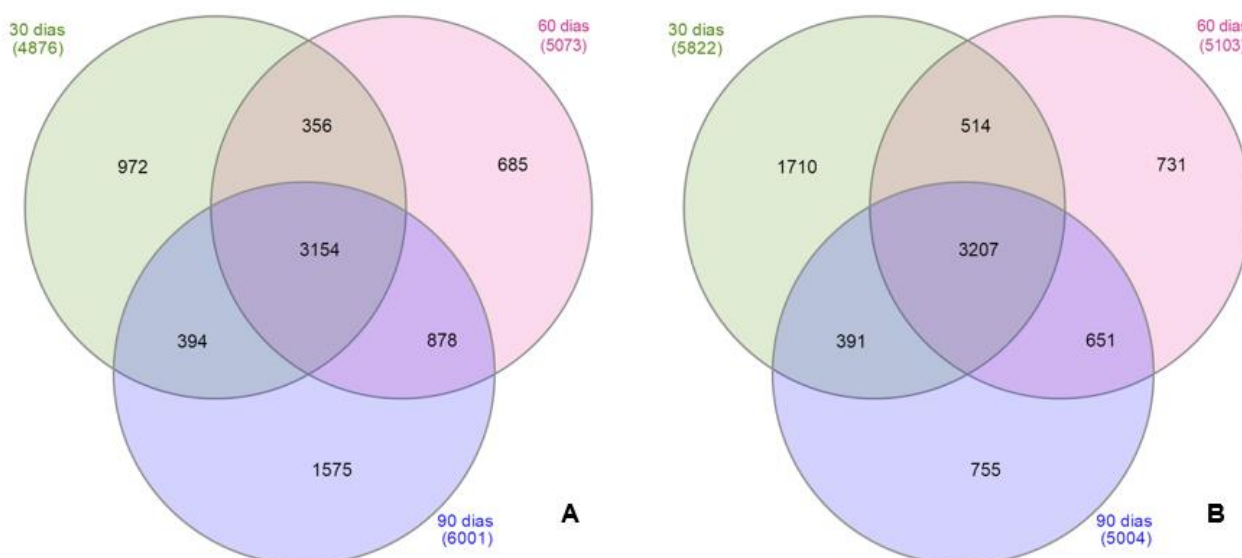
Aos 60 dias de estresse, o número de genes induzidos na cultivar sensível (5103) é menor do que o número de genes induzidos em 30 dias (5822), entretanto esse número ainda é maior do que os GDEs da cultivar tolerante, a qual teve um aumento de 4877 aos 30 dias para 5073 aos 60 dias. Essa cultivar apresentou também um aumento de genes envolvidos em regulação enzimática e resposta a estímulos, ultrapassando assim a cultivar sensível nestas categorias (Figura 18 B).

Aos 90 dias de déficit hídrico, o número de genes diferencialmente expressos induzidos na cultivar tolerante ultrapassa o número dos mesmos para a cultivar sensível (6001 na tolerante e 5004 para a sensível). Além disso, na categoria regulação enzimática ocorre uma queda na expressão de GDEs da cultivar tolerante e um aumento para a cultivar sensível (Figura 18 C).



**Figura 18.** Categorização dos genes diferencialmente expressos (GDEs) induzidos nas duas cultivares estudadas em cada um dos tempos amostrais. A- 30 dias; B-60 dias e C- 90 dias de déficit hídrico.

A relação entre os genes diferencialmente expressos induzidos que são comuns ou exclusivos dentro de cada cultivar, nas três diferentes épocas amostrais (30, 60 e 90 dias), estão representados no Diagrama de Venn (Figura 19).



**Figura 19.** Diagramas de Venn mostrando o número de genes que foram induzidos nas cultivares tolerante (A) e sensível (B) nos tempos de 30, 60 e 90 dias após estresse. Os diagramas mostram o número de genes induzidos em cada tempo e o número de genes que são expressos em comum ou são exclusivos de cada tempo.

#### 4.3.6.1 Análise de agrupamentos de perfis de expressão diferencial para a cultivar tolerante (SP81-3250)

Para a análise de clustering foram selecionados os genes diferencialmente expressos comuns entre as três diferentes épocas amostrais. Esse número de genes foi de 6361, sendo 3154 induzidos na cultivar tolerante em relação a cultivar sensível e 3207 genes induzidos na cultivar sensível em relação à cultivar tolerante (Figura 19).

A análise agrupou os genes em quatro clusters (de acordo com a razão de  $\log_2FC$  de uma cultivar em relação a outra): dois contendo genes induzidos na cultivar sensível (o primeiro com valores de  $\log_2FC$  de 0 a 6,0 e o segundo com valores entre

6,0 e 14,0) e dois contendo genes induzidos na cultivar tolerante (valores de  $\log_2FC$  de 0 a -6,0 no primeiro cluster e -6,0 a -14,0 no segundo).

Para cada um dos clusters gerados foram analisados os 20 genes com os maiores valores de  $\log_2FC$  que puderam ser anotados através das metodologias descritas no Material e Métodos.

No cluster 1, onde agruparam-se os genes induzidos na cultivar tolerante (Tabela 9) em relação à cultivar sensível, foi possível identificar uma proteína não caracterizada (LOC100845726- *Brachypodium distachyon*), composta pelo domínio proteico conservado Peptidase\_S10, serine carboxypeptidase (Pfam00450).

As peptidases (proteases, proteinases ou enzimas proteolíticas) são enzimas que quebram ligações peptídicas, a ligação covalente mais importante que une os aminoácidos para formar peptídeos e proteínas. Nesse processo, uma molécula de água é utilizada e por isso, as peptidases são classificadas como hidrolases. Essas proteínas, identificadas a partir de tomate silvestre e *Arabidopsis*, catalisam a transacetilação no metabolismo secundário em plantas (SHIRLEY; CHAPPLE, 2003), e além do papel nutricional, onde hidrolisam grandes cadeias polipeptídicas em pequenas moléculas que a célula possa absorver, estão envolvidas também em outros processos biológicos, como formação e germinação de esporos e ciclo celular (FEDATTO, 2004).

Tais proteinases já foram identificadas em folhas de tomate em resposta a ferimentos (MOURA; BERGEY; RYAN, 2001), e o genoma de *Arabidopsis* possui cerca de 51 proteínas anotadas para as enzimas de carboxipeptidase (FRASER et al., 2007).

Liu et al. (2008) também encontraram tais enzimas em folhas, raízes, caules e espiga de arroz sob estresse biótico e oxidativo. Nesse estudo, plantas transgênicas de *Arabidopsis* com expressão constitutiva de um gene para essa enzima (OsBISCPL1) foram geradas e os ensaios de resistência a doenças indicou que as plantas transgênicas mostraram maior resistência à doença, em comparação às plantas do tipo selvagem. Além disso, as plantas modificadas, também mostraram uma tolerância aumentada ao estresse oxidativo.

Os resultados em nosso trabalho sugerem que as enzimas carboxipeptidases podem estar envolvidas também no processo de tolerância ao déficit hídrico,

desencadeando uma cascata de resposta na ausência de água, o que favorece o melhor desempenho da planta. O comportamento deste gene ao longo do tempo pode ser observado no gráfico da Figura 20. Notamos que ele diminui sua expressão gradativamente até os 60 dias e depois se mantém constante.

Identificou-se ainda na cultivar tolerante a enzima, 4-cumarato: coenzima A ligase (4CL) (Tabela 8), uma enzima chave nas vias metabólicas para fenilpropanóides e biossíntese dos flavonóides (GUI; SHEN; LI, 2011). Estas vias fenilpropanóides geram diferentes classes de compostos naturais (metabólitos secundários) com funções essenciais no desenvolvimento das plantas e das interações ambientais, tais como lignina para suporte estrutural, flavonas e flavonóis para proteção UV, antocianinas, chalconas e auronas como pigmentos para a atração de polinizadores e os distribuidores de sementes e isoflavonas para defesa de patógenos (HAMBERGER; HAHLBROCK, 2004).

As plantas necessitam de três enzimas para a reação sequencial da via geral de fenilpropanóides: fenilalanina amônia-liase (PAL), cinamato 4-hidroxilase (C4H), e 4-cumarato: CoA-ligase (4CL). Como última enzima da via fenilpropanóide geral, a 4CL regula a ativação do ácido cinâmico e seus derivados para seus tioésteres correspondentes. Estes tioésteres são intermediários para vias centrais específicas que conduzem à síntese de flavonóides e lignina (HAHLBROCK; SCHEEL, 1989; DIXON; PAIVA, 1995; DOUGLAS, 1996), os quais controlam as várias funções fisiológicas em plantas e facilitam a adaptação às perturbações ambientais (DIXON; PAIVA, 1995).

A lignina é responsável pela proteção dos polissacarídeos da parede celular contra a degradação microbiana e proporciona suporte mecânico para toda a planta. Seu teor varia de acordo com o tipo de célula em estudo (VANHOLME et al, 2010; NEUTELINGS et al, 2011).

Em um estudo com *Hibiscus cannabiuns* submetido a períodos de restrição hídrica, ocorreu maior transcrição de um gene 4CL nos tempos iniciais (7 dias) após a retenção da água, e, em seguida, diminuiu nos tempos posteriores (CHOWDHURY et al., 2013).

As alterações no teor de lignina e a expressão do gene podem ser dependentes dos tecidos e o tempo de exposição à seca (VINCENT et al., 2005; FAN et al., 2006;



YANG et al., 2006; LEE et al., 2007; YOSHIMURA et al., 2008). Além disso, a seca pode causar redução do crescimento com deposição de lignina e aumento da expressão de genes de lignina, tais como *PAL*, *C3H*, *4CL*, *CCoAOMT*, *COMT*, *CAD* e *POD*. O aumento da deposição de lignina prejudica a extensibilidade da parede celular, diminui a expansão da parede da célula e reduz o crescimento, o que poderia ser benéfico para a disponibilidade de água (FAN et al., 2006).

Em nosso trabalho, a expressão deste gene diminui gradativamente até os 90 dias de estresse (Figura 20).

No cluster 1, encontrou-se um gene induzido na cultivar tolerante com pico de expressão aos 30 dias de estresse com diminuição aos 60 e 90 dias (Figura 20). Esse gene foi anotado como *pathogenesis-related protein PRB1-3-like*, que possui o domínio *Cysteine-rich secretory protein family (CRISPs)*, um grupo de glicoproteínas, ainda pouco conhecido em plantas. Essas glicoproteínas são ricas no aminoácido cisteína em sua estrutura principal. Os resíduos de cisteína são oxidados para cistina em proteínas, com a formação de ligações dissulfeto, que desempenha um papel no enovelamento de proteínas e a estabilização da estrutura terciária (SEVIER; KAISER, 2002). Em nosso trabalho, supõe-se que essas glicoproteínas, podem estar atuando durante o estresse celular prevenindo a formação de proteínas com conformação tridimensional incorreta durante o déficit hídrico.

A cultivar tolerante apresentou um gene com anotação de *cullin-1-like* com pico de expressão aos 60 dias de estresse e diminuição após esse período (Figura 20).

As proteínas *cullin* funcionam como suportes moleculares que têm papéis cruciais na modificação pós- traducional de proteínas celulares que envolvem ubiquitina. São encontradas em *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana* e leveduras. Estudos genéticos em vários organismos modelo ajudaram a desvendar uma série de funções fisiológicas associadas com as proteínas *cullin*, e podem assim, ter um impacto sobre uma multiplicidade de processos biológicos, incluindo crescimento celular, desenvolvimento, transdução de sinal, controle transcricional, integridade genômica e supressão de tumores (SARIKAS; HARTMANN; PAN, 2011).

Em *Arabidopsis*, essas proteínas regulam a sinalização hormonal, as respostas à luz, ritmos circadianos e fotomorfogênese (HOTTON, CALLIS, 2008). O papel da

proteína *cullin* tipo 1 (CUL1) , está relacionado ao controle do ciclo celular, e é compreendida em considerável detalhe (SARIKAS; HARTMANN; PAN, 2011).

Identificamos ainda nas plantas tolerantes um gene induzido que codifica uma aquaporina (AQP) com baixa expressão aos 30 e 90 dias e aumento aos 60 dias (Figura 20). Sabe-se que o crescimento e desenvolvimento das plantas são dependentes de regulação do movimento da água na planta. A difusão de água através das membranas celulares é facilitada por aquaporinas que fornecem os meios para modificar de forma rápida e reversível a permeabilidade à água. Isto é feito mudando a densidade e atividade de aquaporina na membrana, incluindo modificações pós-traducionais e interação com proteínas que atuam sobre o seu tráfego. Dessa forma, elas podem agir em conjunto com a regulação dos estômatos, bem como na condutividade de raízes e folhas (CHAUMONT; TYERMAN, 2014; JARZYNIAK; JASINSKI, 2014).

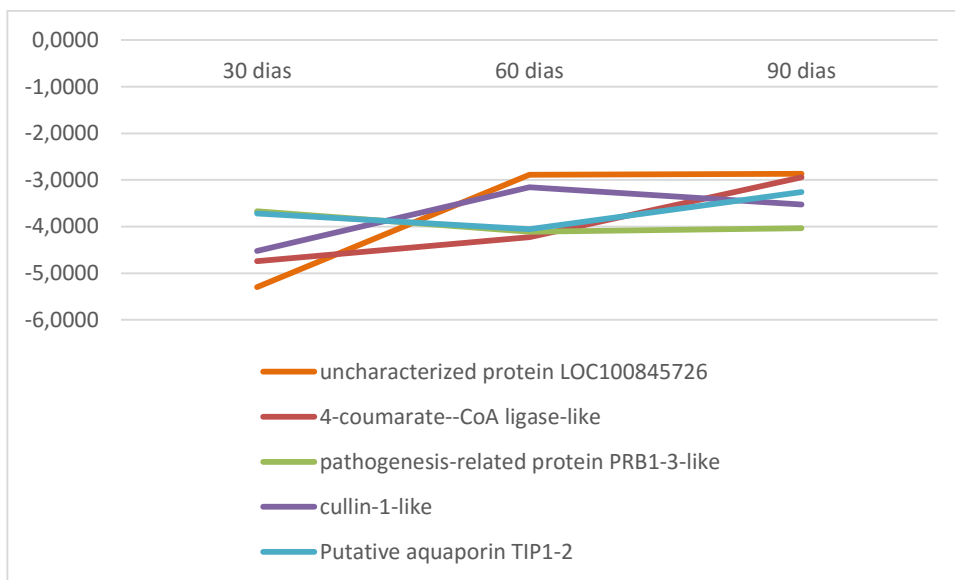
Duas visões opostas existem hoje sobre as funções de AQPs em plantas sob estresse hídrico. A primeira é que o aumento da AQP pode fornecer níveis adicionais para a planta lidar com estresse hídrico, o que se verifica com a indução ou ativação de aquaporinas durante estresse hídrico. A segunda visão é que as plantas tentam evitar a perda excessiva de água por queda na expressão de AQPs durante a desidratação (HACHEZ; ZELAZNY; CHAUMONT, 2006).

Alexandersson et al. (2005) mostraram diferentes padrões de expressão de genes AQP em *Arabidopsis* sob estresse hídrico. A tendência geral foi uma sub-regulação de três genes relacionados com subfamílias das AQP. Curiosamente, estes três genes são modestamente expressos em raízes e altamente expressos em folhas e / ou flores.

Um estudo com sorgo identificou um total de 41 genes de AQP, que foram classificadas em quatro subfamílias (PIPs, TIPs, NIPs e SIPs), dispersos de maneira desigual em todo o genoma. Essas plantas foram submetidas à déficit de ácido abscísico (ABA), estresse hídrico, frio e calor, e a análise de transcrição destes genes em diferentes tecidos revelou que podem desempenhar um papel importante no crescimento e desenvolvimento da planta (REDDY et al., 2015).

Em cana-de-açúcar, 42 isoformas distintas de aquaporinas foram expressas a partir de raízes de cana com genótipo tolerante e sensível à seca. Nessas isoformas

foi possível identificar as quatro subfamílias de aquaporinas. Três genes foram validados via qRT-PCR e revelaram que a resposta ao estresse é específico de cada genótipo (SILVA et al., 2013).



**Figura 20.** Comportamento de cinco genes induzidos na cultivar tolerante em relação à cultivar sensível (Cluster 1) ao longo do tempo (30, 60 e 90 dias).

**Tabela 8.** Descrição dos genes encontrados no Cluster 1 após blastx contra banco de dados não redundante (“nr”)

CLUSTER 1 – Cultivar Tolerante								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g35613	687	<i>hypothetical protein SORBIDRAFT_01g003115 [Sorghum bicolor]</i>	1,00E-76	62%	XP_002466185.1	-5,7939	-2,4094	-2,4077
c0_g103113	271	<i>PREDICTED: uncharacterized protein LOC100845726 [Brachypodium distachyon]</i>	6,00E-05	89%	XP_003557772.2	-5,2990	-2,8906	-2,8670
c0_g13807	447	<i>hypothetical protein SORBIDRAFT_01g022470 [Sorghum bicolor]</i>	2,00E-06	97%	XP_002467277.1	-4,8455	-3,8674	-3,0834
c0_g28316	320	<i>PREDICTED: 4-coumarate—CoA ligase-like 1-like [Setaria italica]</i>	5,00E-51	81%	XP_004985664.1	-4,7440	-4,2272	-2,9402
c0_g105192	697	Nenhuma similaridade significativa encontrada	-	-	-	-4,7326	-4,3727	-2,9445
c0_g88019	374	<i>hypothetical protein [Stenotrophomonas maltophilia]</i>	3.8	33%	WP_032127317.1	-4,6577	-4,4468	-2,8587
c0_g20954	1306	Nenhuma similaridade significativa encontrada	-	-	-	-4,6533	-3,9895	-2,6449
c0_g132246	991	<i>hypothetical protein [Yersinia rohdei]</i>	2.0	30%	WP_032817318.1	-4,6042	-3,9610	-3,2596
c0_g19700	302	<i>Phospholipase D LcsSicTox-beta1C1(PLD)</i>	5,00E-10	54%	C0JB40.1	-4,5911	-3,7599	-3,1096
c0_g90410	1105	Nenhuma similaridade significativa encontrada	-	-	-	-4,5639	-3,3830	-3,0874
c0_g71234	446	Nenhuma similaridade significativa encontrada	-	-	-	-4,5544	-3,9854	-2,6972
c0_g116365	522	Nenhuma similaridade significativa encontrada	-	-	-	-4,5469	-4,1775	-2,4155
c0_g33047	333	<i>hypothetical protein [Shewanella baltica]</i>	8.3	44%	WP_012088333.1	-4,5407	-3,7649	-3,0135
c0_g96208	377	<i>hypothetical protein SORBIDRAFT_04g013083 [Sorghum bicolor]</i>	7,00E-06	60%	XP_002453750.1	-4,5072	-4,2066	-2,3839
c0_g45025	315	<i>hypothetical protein [Enterococcus moraviensis]</i>	2.7	33%	WP_010766294.1	-4,4635	-3,8302	-2,8029
c0_g75701	773	Nenhuma similaridade significativa encontrada	-	-	-	-4,4463	-4,1976	-3,3673
c0_g109367	690	<i>uncharacterized protein LOC100191850 [Zea mays]</i>	6,00E-40	81%	NP_001130746.1	-4,4021	-3,8721	-2,6312

Continuação								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g123719	2425	<i>OSJNBb0062H02.17 [Oryza sativa Japonica Group]</i>	0.0	58%	CAD40278.2	-4,4015	-3,6691	-2,9725
c0_g63207	2708	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-4,3721	-3,5565	-2,7979
c0_g8749	859	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-4,3574	-3,6517	-3,0228
c0_g93284	716	<i>hypothetical protein SORBIDRAFT_03g046922 [Sorghum bicolor]</i>	1,00E-05	84%	XP_002459158.1	-4,3506	-4,6297	-2,4984
c0_g35220	315	<i>hypothetical protein B456_012G151700 [Gossypium raimondii]</i>	3,00E-15	43%	KJB77699.1	-4,1599	-5,1735	-2,1312
c0_g85905	373	<i>PREDICTED: PQ-loop repeat-containing protein 1 isoform X3 [Mandrillus leucophaeus]</i>	5.3	48%	XP_011831262.1	-4,0087	-4,0236	-3,2090
c0_g69234	1684	<i>hypothetical protein SORBIDRAFT_03g026710 [Sorghum bicolor]</i>	8,00E-71	99%	XP_002458083.1	-3,9873	-4,1435	-3,4923
c0_g7644	659	<i>hypothetical protein LOC_Os12g30780 [Oryza sativa Japonica Group]</i>	1,00E-11	57%	ABA98304.1	-3,8948	-4,1003	-2,7597
c0_g120750	800	<i>Nenhuma similaridade significativa encontrada</i>				-3,8692	-4,1127	-3,8220
c0_g111898	1983	<i>PREDICTED: uncharacterized protein LOC101775145 [Setaria italica]</i>	0.0	88%	XP_004967061.1	-3,8475	-3,9178	-4,1165
c0_g109175	430	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-3,8230	-4,6884	-3,5324
c0_g109313	721	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-3,8209	-4,1895	-3,4315
c0_g64038	520	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-3,8208	-4,1124	-3,1273
c0_g107652	699	<i>PREDICTED: putative disease resistance protein RGA4-like [Setaria italica]</i>	1,00E-31	56%	XP_004967143.1	-3,8158	-3,8758	-3,9991
c0_g88488	465	<i>retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]</i>	4,00E-08	31%	AAX95863.1	-3,7669	-3,7733	-4,3134
c0_g64412	2507	<i>PREDICTED: uncharacterized protein LOC101755730 [Setaria italica]</i>	0.0	51%	XP_004954645.1	-3,7444	-3,8143	-4,3694

Continuação								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g94137	201	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-3,7405	-4,2204	-2,6682
c0_g96212	1555	<i>PREDICTED: uncharacterized protein LOC101766936 [Setaria italica]</i>	3,00E-145	58%	XP_004967999.1	-3,7350	-4,5958	-3,3299
c0_g113944	480	<i>TPA: hypothetical protein ZEAMMB73_658394 [Zea mays]</i>	5,00E-12	85%	DAA56665.1	-3,7248	-3,9464	-3,1231
c0_g18041	526	<i>Putative aquaporin TIP1-2 [Aegilops tauschii]</i>	1,00E-04	93%	EMT32480.1	-3,7203	-4,0583	-3,2579
c0_g36339	632	<i>hypothetical protein SORBIDRAFT_06g015900 [Sorghum bicolor]</i>	3,00E-63	87%	XP_002446423.1	-3,6844	-3,8186	-2,9227
c0_g108784	942	<i>putative stripe rust resistance protein Yr10 [Sorghum bicolor]</i>	5,00E-134	82%	AAM94294.1	-3,6840	-3,9583	-4,0414
c0_g53720	477	<i>PREDICTED: pathogenesis-related protein PRB1-3-like [Zea mays]</i>	5,00E-10	88%	XP_008651988.1	-3,6737	-4,1102	-4,0332
c0_g111898	1983	<i>PREDICTED: uncharacterized protein LOC101775145 [Setaria italica]</i>	0.0	88%	XP_004967061.1	-3,8475	-3,9178	-4,1165
c0_g107652	699	<i>PREDICTED: putative disease resistance protein RGA4-like [Setaria italica]</i>	1,00E-31	56%	XP_004967143.1	-3,8158	-3,8758	-3,9991
c0_g88488	465	<i>retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]</i>	4,00E-08	31%	AAX95863.1	-3,7669	-3,7733	-4,3134
c0_g64412	2507	<i>PREDICTED: uncharacterized protein LOC101755730 [Setaria italica]</i>	0.0	51%	XP_004954645.1	-3,7444	-3,8143	-4,3694
c0_g108784	942	<i>putative stripe rust resistance protein Yr10 [Sorghum bicolor]</i>	5,00E-134	82%	AAM94294.1	-3,6840	-3,9583	-4,0414
c0_g133833	1020	<i>hypothetical protein SORBIDRAFT_06g002845 [Sorghum bicolor]</i>	7,00E-22	56%	XP_002447531.1	-3,6555	-3,7185	-3,9789
c0_g115870	322	<i>Nenhuma similaridade significativa encontrada</i>	-	-	-	-3,6273	-4,0117	-4,1848
c0_g87629	2985	<i>unnamed protein product [Oryza sativa Japonica Group]</i>	0.0	40%	BAC15618.2	-3,5378	-3,8456	-3,9473
c0_g28427	237	<i>ligand-gated channel protein [Rhodobacter sphaeroides]</i>	6.2	33%	WP_044249893.1	-3,4178	-4,2253	-4,3831

Continuação								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g60869	270	Nenhuma similaridade significativa encontrada	-	-	-	-3,3746	-4,0896	-4,4724
c0_g16579	949	hypothetical protein SORBIDRAFT_09g009990 [Sorghum bicolor]	2,00E-119	64%	XP_002440867.1	-3,3414	-3,8630	-3,9708
c0_g40033	463	Nenhuma similaridade significativa encontrada	-	-	-	-3,3382	-3,3434	-3,8018
c0_g133170	914	Nenhuma similaridade significativa encontrada	-	-	-	-3,2666	-3,5603	-4,2722
c0_g22295	323	hypothetical protein SORBIDRAFT_07g028100 [Sorghum bicolor]	7,00E-41	70%	XP_002445924.1	-3,2220	-3,3655	-3,9035
c0_g52129	952	Nenhuma similaridade significativa encontrada	-	-	-	-3,2193	-3,6634	-4,3826
c0_g58819	538	PREDICTED: ankyrin repeat, PH and SEC7 domain containing protein secG-like [Setaria italica]	0.009	57%	XP_004964171.1	-3,2175	-3,4154	-4,0523
c0_g108342	430	LysR family transcriptional regulator [Salinisphaera hydrothermalis]	1.4	38%	WP_037337706.1	-3,2106	-3,5215	-4,4403
c0_g119650	2454	hypothetical protein SORBIDRAFT_06g010506 [Sorghum bicolor]	8,00E-61	43%	XP_002447633.1	-3,1811	-3,5760	-4,3856
c0_g13138	978	Acetyltransferase, GNAT [Bacillus cereus m1293]	0.84	33%	EEK45605.1	-3,1401	-3,2941	-4,5908
c0_g2664	1380	hypothetical protein PISMIDRAFT_13860 [Pisolithus microcarpus 441]	2.1	24%	KIK19170.1	-3,1253	-3,5856	-3,8268
c0_g113354	1295	Ascorbate peroxidase [Medicago truncatula]	0.015	79%	XP_003612205.1	-6,1074	-2,3993	-3,5181
c0_g46753	331	TPA: hypothetical protein ZEAMMB73_957883 [Zea mays]	0.024	63%	DAA60311.1	-5,8134	-2,9777	-2,9846
c0_g106139	803	Nenhuma similaridade significativa encontrada	-	-	-	-5,7566	-2,3089	-3,7091
c0_g80365	581	Nenhuma similaridade significativa encontrada	-	-	-	-5,1230	-2,5012	-2,9431
c0_g33807	756	hypothetical protein OsI_09298 [Oryza sativa Indica Group]	9,00E-95	79%	EAY87878.1	-5,0825	-2,0987	-2,8890
c0_g11694	450	hypothetical protein [Nocardiopsis chromatogenes]	3.4	35%	WP_040725772.1	-5,0038	-2,5481	-3,7964

Continuação								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g15969	485	Nenhuma similaridade significativa encontrada	-	-	-	-4,9306	-2,3922	-4,5939
c0_g40218	831	hypothetical protein [Sphingomonas sp. DC-6]	9.1	39%	WP_030091701.1	-4,8366	-2,1877	-2,7548
c0_g54148	1474	hypothetical protein SORBIDRAFT_03g002980 [Sorghum bicolor]	0.0	94%	XP_002455011.1	-4,8222	-2,6184	-3,8020
c0_g112490	1236	helicase-like protein [Oryza sativa Japonica Group]	1,00E-159	75%	BAD68127.1	-4,7678	-3,1403	-3,2466
c0_g92220	322	Nenhuma similaridade significativa encontrada	-	-	-	-4,7362	-3,1536	-3,5467
c0_g107649	422	hypothetical protein SORBIDRAFT_05g003620 [Sorghum bicolor]	6,00E-05	73%	XP_002450316.1	-4,6564	-2,8209	-3,8348
c0_g19165	461	hypothetical protein SORBIDRAFT_02g001620 [Sorghum bicolor]	5,00E-27	83%	XP_002461366.1	-4,5960	-2,4065	-2,7983
c0_g50821	1150	PREDICTED: uncharacterized protein LOC101777680 [Setaria italica]	2,00E-84	83%	XP_004980553.1	-4,5914	-3,2779	-4,1191
c0_g9400	286	peptide ABC transporter ATP-binding protein [Mesorhizobium sp. SOD10]	4.3	43%	WP_040994315.1	-4,5807	-2,7937	-3,6389
c0_g30089	694	Nenhuma similaridade significativa encontrada	-	-	-	-4,5691	-3,1714	-3,1913
c0_g103739	216	hypothetical protein [Enterococcus avium]	7.3	35%	WP_016178288.1	-4,5575	-2,7809	-3,4066
c0_g22750	672	hypothetical protein PRUPE_ppa021335mg [Prunus persica]	6,00E-04	81%	XP_007203404.1	-4,5495	-3,0301	-3,3254
c0_g82002	995	PREDICTED: cullin-1-like [Brachypodium distachyon]	2,00E-08	71%	XP_003568930.1	-4,5207	-3,1528	-3,5232
c0_g14616	410	Nenhuma similaridade significativa encontrada	-	-	-	-4,5170	-2,3528	-3,5973



No cluster 2 da cultivar tolerante foi identificado uma proteína de membrana de ligação de esteróides (*membrane steroid-binding protein 1* – MSBP1) (Tabela 10), com pico de expressão aos 30 e 90 dias de estresse (Figura 22).

Essa proteína que também foi identificada e caracterizada em *Arabidopsis thaliana*, funciona como um regulador negativo de alongamento e expansão celular (YANG; XU; XUE, 2005) e também da sinalização de brassinosteróides (BRs) (SHI et al., 2011).

Yang, Xu e Xue (2005), em seu experimento verificaram que as plantas transgênicas com baixa expressão de MSBP1 demonstraram fenótipos de hipocótilo longo, reduzindo assim a capacidade de ligação dos esteróides. Por outro lado, plantas transgênicas superexpressando MSBP1 mostraram um aumento na capacidade de ligação de esteróides de membrana, e exibiram um fenótipo de hipocótilo curto.

O alongamento celular reduzido de MSBP1, foi relacionado com a expressão alterada de genes envolvidos no alongamento de células, tais como as expansinas, indicando que uma maior quantidade de MSBP1 afeta vias para o alongamento celular. Além disso, a expressão de MSBP1 em hipocótilos é suprimida pela escuridão e ativada pela luz, o que sugere que MSBP1 atua como um regulador negativo do alongamento das células. (YANG; XU; XUE, 2005).

Também no cluster 2, identificamos um gene com pico de expressão aos 30 e 60 dias de estresse (Figura 22), com anotação para uma proteína não caracterizada (*LOC101772414*). Essa proteína apresenta a região MuDR family transposase, uma região encontrada em proteínas vegetais. O elemento MuDr regula a transposição de uma família de elementos transponíveis Mu (*Mutator*) que estão presentes no genoma do milho em grande número de cópias (EMBRAPA, 2010).

Os elementos transponíveis podem ser divididos em duas grandes classes, com base no mecanismo utilizado para a transposição, e na organização estrutural do elemento. Os elementos MuDR se incluem na classe II que compreende os elementos que se transpõem diretamente de DNA a DNA, chamados de transposons e usam a transposase, uma enzima codificada por eles. Esses elementos de classe II podem se excisar precisamente do sítio original e se inserir em um novo local do genoma,

podendo, até mesmo, reverter mutações originalmente causadas pela sua inserção (DEPRÁ, 2005).

De acordo com Cruz (2012), através do projeto SUCEST (Sugarcane EST Project), dentro da caracterização dos genes hipotéticos da cana, as famílias mais representadas dos elementos de classe II foram MuDr, Ac e Tam3. Tais elementos foram encontrados expressos em diferentes tecidos e ao longo de diferentes estágios de crescimento da planta.

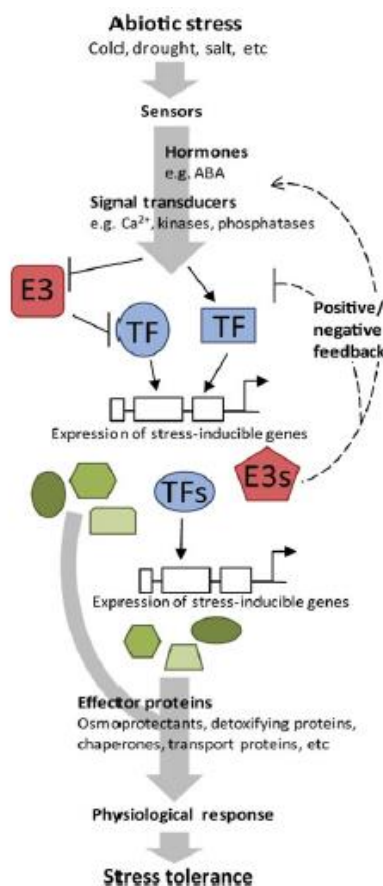
Ainda no cluster 2 identificamos um gene com descrição *E3 SUMO-protein ligase*. Essas proteínas SUMO (Small ubiquitin-related modifier), ligam-se de forma covalente e modificam outras proteínas que participam de diversos processos celulares, incluindo regulação da transcrição, o transporte nuclear, a manutenção da integridade do genoma, e a transdução de sinal (JOHNSON, 2004).

Em plantas, as proteínas SUMO estão intimamente relacionadas à resposta a estímulos ambientais, incluindo o calor, o frio, a seca, salinidade, e ainda, a modulação de ácido abscísico e outros hormônios, bem como a homeostase de nutrientes (ZHANG et al., 2013).

Em *Arabidopsis*, a seca induz um acúmulo das proteínas SUMO, um processo parcialmente dependente da ligase E3 - SIZ1. A SIZ1 parece agir positivamente na tolerância à seca mostrando sensibilidade desde curto a longos prazos de desidratação (CASTRO et al., 2012).

As E3 ligases estão envolvidas com a via de ubiquitinação e portanto, controlam numerosos processos celulares. A maioria delas, identificadas até o momento, parecem facilitar as respostas a estímulos ambientais modulando a abundância dos principais fatores de transcrição (LYZENGA; STONE, 2012).

Com a identificação de um número crescente de E3 ligases que regulam as respostas a estresses abióticos, os mecanismos de modo de ação das E3 estão se tornando mais definidos (Figura 21). Essas ligases podem funcionar por meio da supressão da via de sinalização de estresse durante condições favoráveis de crescimento, ou atenuando-a quando as condições melhoram, permitindo o crescimento. As E3 ligases também podem exercer função dentro de um ciclo de feedback positivo para aumentar a sinalização ao estresse (Figura 21).



**Figura 21.** Regulação do estresse abiótico via sinalização por E3 ligases. A planta percebe os sinais de estresse através de sensores (desconhecido) e o sinal é enviado via hormônios vegetais, mensageiros secundários e reguladores da transcrição. A expressão induzida dos genes de estresse é facilitada por fatores de transcrição (TF). As E3 ligases tendem a regular os componentes da via de sinalização, principalmente os TFs. Na ausência de um sinal, as ligases podem suprimir a via de sinalização, por exemplo, promovendo a degradação de um TF. E3 ligases pode exercer uma função dentro de um mecanismo de *feedback* para melhorar ou atenuar o sinal do estresse. FONTE: LYZENGA; STONE, 2012.

Catala et al. (2007) em seus experimentos com *Arabidopsis* demonstraram que mutantes que não possuíam a E3 - SIZ1, quando submetidos ao déficit hídrico, apresentaram menor tolerância à seca e menor estatura por conta da menor expressão de genes envolvidos na biossíntese e sinalização de brassinosteróides. No genoma de *Arabidopsis*, 1700 genes foram induzidos pela seca, sendo que o SIZ1 é responsável por mediar a expressão de 300 deles. Tais resultados reforçam que a E3 ligase desempenha papel importante na resposta à seca através da regulação da expressão gênica.

Em cana de açúcar, o longo período de estresse pode ter induzido um acúmulo dessas proteínas e conseqüentemente tolerância ao déficit hídrico. De acordo com a Figura 22 é possível verificar que a expressão desse gene aumentou até os 60 dias de estresse, mantendo-se constante até os 90 dias de déficit hídrico.

No cluster 2 da cultivar tolerante foi identificado um gene com anotação para a proteína argonauta. Várias classes de proteínas argonautas e pequenos RNAs são encontradas em todos os organismos eucariotos e têm funções importantes em processos diversos como o desenvolvimento embrionário, a diferenciação de células e silenciamento de transposons (HOCK; MEISTER, 2008).

Pequenos RNAs não codantes emergiram como moléculas chave na regulação de vários processos biológicos. Em plantas, microRNAs (miRNAs) e RNAs de interferência (iRNAs) representam a grande maioria da população de pequenos RNAs. Estes são processados pelas enzimas DICER-LIKE (DCL1) e interagem com as enzimas Argonautas (AGO), efetuando o silenciamento gênico (PINOTI, 2012).

Para manter o crescimento e o desenvolvimento normal de uma planta, a expressão do gene deve estar sob rigorosa vigilância. Uma das vias regulatórias pós-transcricional envolve pequenos RNAs (sRNA) que orientam as proteínas argonautas aos seus objetivos específicos através de complementaridade de sequência, que conduz então à clivagem do mRNA ou inibição da tradução. (SHAO et al, 2014).

As enzimas argonautas desempenham papéis cruciais no silenciamento de RNA. São proteínas altamente conservadas e são definidas pela presença dos domínios PAZ e PIWI. Em *Arabidopsis*, a família AGO compreende dez membros, e é provável que diversas funções de silenciamento de RNA estejam relacionadas com a variação nos membros dessas famílias (QIAN et al, 2011).

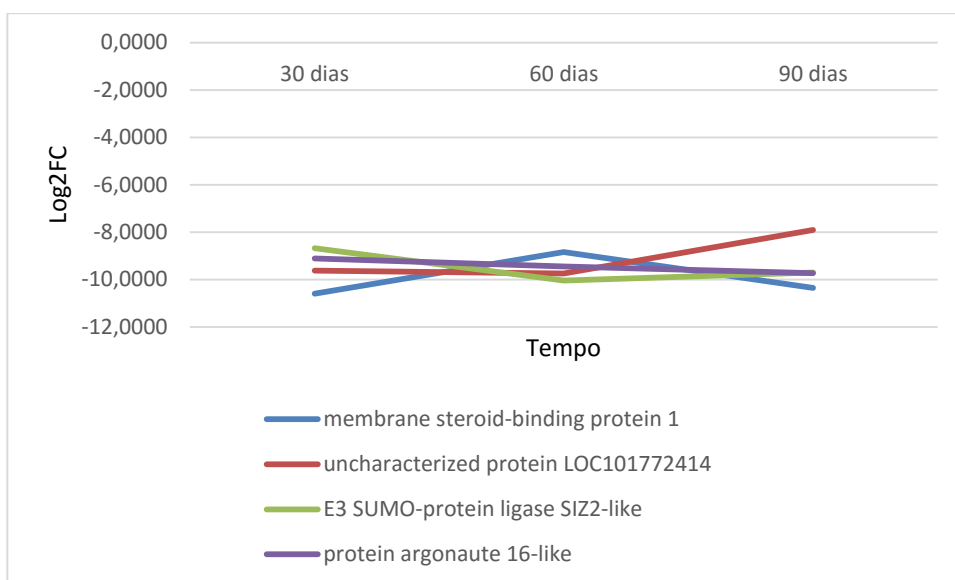
Yigit et al (2006) relataram que o arroz possui o maior número de argonautas entre as plantas, quase dobrando o número relatado em *Arabidopsis*.

Qian e colaboradores (2011) em seu experimento, encontraram quase o mesmo número de membros de AGO em milho e supõem que essas famílias de proteínas de milho, arroz e *Arabidopsis* partilham grandes semelhanças.

Em um estudo sobre estresse hídrico e estresse salino com *Panicum virgatum*, foi observada uma alta expressão de miRNAs conservados e de alguns miRNAs recém identificados, induzidos pela seca e salinidade (XIE et al., 2014).

Em cana de açúcar, devido a ausência da sequência genômica, a identificação de miRNAs é muito complicada. Seus precursores são altamente instáveis, tornando a sua detecção em ESTs de cana muito difícil (VETTORE et al., 2003).

Em uma revisão sobre expressão de miRNAs em cana-de-açúcar sob déficit hídrico (GENTILE et al., 2015) afirma-se que o padrão de expressão destes RNAs é variável, dependendo das espécies, do tipo de tratamento, do tecido envolvido, da cultivar e condições de crescimento (campo, casa de vegetação, por exemplo). Acredita-se que os miRNAs são regulados por uma via complexa, e que as plantas podem ajustar a expressão dos mesmos de diversas maneiras, de modo a lidar com diferentes fases e intensidade do estresse hídrico.



**Figura 22.** Comportamento de quatro genes induzidos na cultivar tolerante em relação à cultivar sensível (Cluster 2) ao longo do tempo (30, 60 e 90 dias).

**Tabela 9.** Descrição dos genes encontrados no Cluster 2 após blastx contra banco de dados não redundante (“nr”)

CLUSTER 2– Cultivar Tolerante								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g34267	266	Nenhuma similaridade significativa encontrada	-	-	-	-11,1969	-7,9270	-10,4007
c0_g4716	1109	hypothetical protein SORBIDRAFT_01g049310 [Sorghum bicolor]	3,00E-13	56%	XP_002468628.1	-11,0807	-9,0885	-11,3632
c0_g123767	1258	uncharacterized protein LOC100384284 [Zea mays]	5,00E-41	89%	NP_001170317.1	-10,7337	-6,7775	-8,0049
c0_g117932	407	pothetical protein, conserved [Leishmania donovani]	3.9	34%	XP_003860095.1	-10,6848	-10,3079	-10,4829
c0_g110486	374	hypothetical protein POPTR_0008s17550g [Populus trichocarpa]	2.0	35%	XP_002312621.2	-10,6131	-7,4389	-8,5454
c0_g35794	1102	membrane steroid-binding protein 1 [Zea mays]	1,00E-45	80%	NP_001150181.1	-10,6042	-8,8465	-10,3494
c0_g103873	683	Nenhuma similaridade significativa encontrada				-10,5365	-8,4416	-7,2568
c0_g85162	334	hypothetical protein SORBIDRAFT_02g004840 [Sorghum bicolor]	0.16	89%	XP_002461567.1	-10,5005	-8,1451	-6,4650
c0_g16445	1959	myb-like DNA-binding domain containing protein [Zea mays]	0.005	87%	NP_001151853.1	-10,3993	-8,0327	-9,8766
c0_g46198	1790	hypothetical protein PAXINDRAFT_130281 [Paxillus involutus ATCC 200175]	1.9	35%	KIJ17846.1	-10,3635	-6,8858	-8,6876
c0_g3833	661	hypothetical protein SORBIDRAFT_09g004440 [Sorghum bicolor]	3,00E-14	98%	XP_002439324.1	-10,2765	-7,7049	-5,7771
c0_g78028	1596	hypothetical protein SORBIDRAFT_02g026130 [Sorghum bicolor]	5,00E-04	60%	XP_002460295.1	-10,2266	-7,3450	-2,7485
c0_g82730	308	Nenhuma similaridade significativa encontrada	-	-	-	-10,2196	-9,0512	-9,6472
c0_g111924	436	Nenhuma similaridade significativa encontrada	-	-	-	-10,1944	-9,2801	-8,4163
c0_g90955	1371	Nenhuma similaridade significativa encontrada	-	-	-	-10,1836	-6,1907	-6,5783
c0_g105191	460	Nenhuma similaridade significativa encontrada	-	-	-	-10,1373	-9,1164	-9,2467

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g64180	1837	PREDICTED: ankyrin-3-like isoform X1 [Setaria italica]	0.0	83%	XP_004978963.1	-10,1348	-8,8773	-9,9306
c0_g56006	240	conserved hypothetical protein [Trichophyton verrucosum HKI 0517]	1.1	37%	XP_003019398.1	-10,1208	-7,0114	-7,6388
c0_g84727	364	fungual zn(2)-Cys(6) binuclear cluster domain-containing protein [Rhizoctonia solani AG-1 IA]	5.3	36%	ELU45075.1	-10,1078	-7,6464	-7,2856
c0_g59082	726	hypothetical protein PILCRDRAFT_819455 [Piloderma croceum F 1598]	9.1	35%	KIM83219.1	-10,0928	-8,2092	-7,4221
c0_g103873	683	Nenhuma similaridade significativa encontrada	-	-	-	-10,5365	-8,4416	-7,2568
c0_g85162	334	hypothetical protein SORBIDRAFT_02g004840 [Sorghum bicolor]	0.16	89%	XP_002461567.1	-10,5005	-8,1451	-6,4650
c0_g3833	661	hypothetical protein SORBIDRAFT_09g004440 [Sorghum bicolor]	3,00E-14	98%	XP_002439324.1	-10,2765	-7,7049	-5,7771
c0_g78028	1596	hypothetical protein SORBIDRAFT_02g026130 [Sorghum bicolor]	5,00E-04	60%	XP_002460295.1	-10,2266	-7,3450	-2,7485
c0_g111924	436	Nenhuma similaridade significativa encontrada	-	-	-	-10,1944	-9,2801	-8,4163
c0_g84727	364	fungual zn(2)-Cys(6) binuclear cluster domain-containing protein [Rhizoctonia solani AG-1 IA]	5.3	36%	ELU45075.1	-10,1078	-7,6464	-7,2856
c0_g59082	726	hypothetical protein PILCRDRAFT_819455 [Piloderma croceum F 1598]	9.1	35%	KIM83219.1	-10,0928	-8,2092	-7,4221
c0_g119451	285	PREDICTED: uncharacterized protein LOC102705538 [Oryza brachyantha]	1,00E-04	81%	XP_006656192.1	-10,0846	-7,8960	-7,6353
c0_g74520	589	Nenhuma similaridade significativa encontrada	-	-	-	-10,0770	-10,1163	-10,2622
c0_g9636	909	hypothetical protein [Sorghum bicolor]	5,00E-06	59%	AAD27576.1	-10,0615	-7,7728	-6,0682
c0_g116001	702	predicted protein [Hordeum vulgare subsp. vulgare]	0.37	29%	BAJ98001.1	-9,9738	-9,6432	-9,0279
c0_g22760	677	Os11g0209000 [Oryza sativa Japonica Group]	2.8	69%	NP_001067475.1	-9,9304	-10,0104	-9,0584
c0_g55261	2102	PREDICTED: uncharacterized protein LOC103636211 [Zea mays]	5,00E-129	49%	XP_008656787.1	-9,9080	-12,9819	-10,1087
c0_g113752	847	BnaA09g39630D [Brassica napus]	0.14	78%	CDY11397.1	-9,8895	-9,7404	-9,3636

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g20750	670	Hypothetical protein CAP_1135 [Chondromyces apiculatus DSM 436]	0.80	29%	EYF00148.1	-9,8661	-7,8434	-7,4957
c0_g111574	1435	hypothetical protein ZEAMMB73_045866 [Zea mays]	8,00E-11	46%	AFW75917.1	-9,7317	-9,4101	-8,5701
c0_g61029	376	hypothetical protein AMTR_s00075p00194500 [Amborella trichopoda]	0.50	32%	ERN04628.1	-9,7310	-9,9226	-8,9115
c0_g120316	671	PREDICTED: uncharacterized protein LOC100193501 isoform X1 [Zea mays]	1,00E-15	97%	XP_008667181.1	-9,6350	-9,9459	-9,7815
c0_g110526	1921	PREDICTED: uncharacterized protein LOC101772414 [Setaria italica]	5,00E-134	76%	XP_004964132.1	-9,6303	-9,7530	-7,9080
c0_g132393	1044	PREDICTED: putative disease resistance RPP13-like protein 3-like isoform X1 [Setaria italica]	4,00E-09	65%	XP_004967978.1	-9,5771	-8,4538	-7,8949
c0_g74520	589	Nenhuma similaridade significativa encontrada	-	-	-	-10,0770	-10,1163	-10,2622
c0_g22760	677	Os11g0209000 [Oryza sativa Japonica Group]	2.8	69%	NP_001067475.1	-9,9304	-10,0104	-9,0584
c0_g55261	2102	PREDICTED: uncharacterized protein LOC103636211 [Zea mays]	5,00E-129	49%	XP_008656787.1	-9,9080	-12,9819	-10,1087
c0_g61029	376	hypothetical protein AMTR_s00075p00194500 [Amborella trichopoda]	0.50	32%	ERN04628.1	-9,7310	-9,9226	-8,9115
c0_g120316	671	PREDICTED: uncharacterized protein LOC100193501 isoform X1 [Zea mays]	1,00E-15	97%	XP_008667181.1	-9,6350	-9,9459	-9,7815
c0_g110526	1921	PREDICTED: uncharacterized protein LOC101772414 [Setaria italica]	5,00E-134	76%	XP_004964132.1	-9,6303	-9,7530	-7,9080
c0_g60502	830	Nenhuma similaridade significativa encontrada	-	-	-	-9,4449	-9,4946	-9,0914
c0_g2552	1046	Nenhuma similaridade significativa encontrada	-	-	-	-9,3646	-9,4472	-9,6216
c0_g41538	1316	Nenhuma similaridade significativa encontrada	-	-	-	-9,2724	-9,7678	-9,4968
c0_g57553	521	hypothetical protein [Paraglaciecola psychrophila]	2.7	36%	WP_007638827.1	-9,2228	-9,3867	-9,2811
c0_g66251	994	D-alanyl-D-alanine carboxypeptidase [Bacillus kribbensis]	2.8	32%	WP_026694004.1	-9,1933	-10,2747	-10,5388



**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g114516	376	hypothetical protein SORBIDRAFT_02g005615 [Sorghum bicolor]	1,00E-19	84%	XP_002459498.1	-9,1610	-10,1026	-9,3074
c0_g115226	756	hypothetical protein SORBIDRAFT_03g011873 [Sorghum bicolor]	7,00E-09	44%	XP_002457703.1	-9,1567	-10,1410	-9,7034
c0_g43638	344	PREDICTED: protein argonaute 16-like [Eucalyptus grandis]	0.006	90%	XP_010035500.1	-9,1127	-9,4484	-9,7392
c0_g116693	874	Nenhuma similaridade significativa encontrada	-	-	-	-9,0115	-10,7546	-10,3652
c0_g10694	1101	PREDICTED: sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like [Leptonychotes weddellii]	3.5	44%	XP_006733545.1	-8,9415	-9,8215	-9,6653
c0_g40659	250	hypothetical protein [Parasitella parasitica]	3.6	35%	CEP08920.1	-8,9147	-9,1907	-9,6862
c0_g72868	2809	PREDICTED: putative disease resistance RPP13-like protein 3 [Zea mays]	0.0	71%	XP_008649401.1	-8,8799	-9,4422	-9,8477
c0_g55616	1599	Nenhuma similaridade significativa encontrada	-	-	-	-8,8741	-8,9963	-8,7596
c0_g119185	2668	Nenhuma similaridade significativa encontrada	-	-	-	-8,8656	-9,2679	-7,3693
c0_g22760	677	Os11g0209000 [Oryza sativa Japonica Group]	2.8	69%	NP_001067475.1	-9,9304	-10,0104	-9,0584
c0_g55261	2102	PREDICTED: uncharacterized protein LOC103636211 [Zea mays]	5,00E-129	49%	XP_008656787.1	-9,9080	-12,9819	-10,1087
c0_g61029	376	hypothetical protein AMTR_s00075p00194500 [Amborella trichopoda]	0.50	32%	ERN04628.1	-9,7310	-9,9226	-8,9115
c0_g120316	671	PREDICTED: uncharacterized protein LOC100193501 isoform X1 [Zea mays]	1,00E-15	97%	XP_008667181.1	-9,6350	-9,9459	-9,7815
c0_g110526	1921	PREDICTED: uncharacterized protein LOC101772414 [Setaria italica]	5,00E-134	76%	XP_004964132.1	-9,6303	-9,7530	-7,9080
c0_g60502	830	Nenhuma similaridade significativa encontrada	-	-	-	-9,4449	-9,4946	-9,0914
c0_g41538	1316	Nenhuma similaridade significativa encontrada	-	-	-	-9,2724	-9,7678	-9,4968
c0_g57553	521	hypothetical protein [Paraglaciecola psychrophila]	2.7	36%	WP_007638827.1	-9,2228	-9,3867	-9,2811
c0_g55616	1599	Nenhuma similaridade significativa encontrada	-	-	-	-8,8741	-8,9963	-8,7596

Continuação								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g114516	376	hypothetical protein SORBIDRAFT_02g005615 [Sorghum bicolor]	1,00E-19	84%	XP_002459498.1	-9,1610	-10,1026	-9,3074
c0_g115226	756	hypothetical protein SORBIDRAFT_03g011873 [Sorghum bicolor]	7,00E-09	44%	XP_002457703.1	-9,1567	-10,1410	-9,7034
c0_g116693	874	Nenhuma similaridade significativa encontrada	-	-	-	-9,0115	-10,7546	-10,3652
c0_g10694	1101	PREDICTED: sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like [Leptonychotes weddellii]	3.5	44%	XP_006733545.1	-8,9415	-9,8215	-9,6653
c0_g119185	2668	Nenhuma similaridade significativa encontrada	-	-	-	-8,8656	-9,2679	-7,3693
c0_g57349	4931	OSJNBb0062H02.17 [Oryza sativa Japonica Group]	0.0	57%	CAD40278.2	-8,8288	-9,5929	-8,6824
c0_g121842	3348	hypothetical protein LOC_Os11g07260 [Oryza sativa Japonica Group]	0.0	66%	AAX93041.1	-8,7156	-9,3504	-9,1723
c0_g121275	803	PREDICTED: trem-like transcript 4 protein [Loxodonta africana]	4.3	33%	XP_010587401.1	-8,6913	-10,3783	-8,9832
c0_g129219	1265	PREDICTED: E3 SUMO-protein ligase SIZ2-like isoform X4 [Setaria italica]	3,00E-23	77%	XP_004977641.1	-8,6802	-10,0484	-9,7016
c0_g27292	413	hypothetical protein SORBIDRAFT_0010s005460 [Sorghum bicolor]	5,00E-59	87%	XP_002489303.1	-8,6784	-9,0822	-8,9988
c0_g108135	1432	hypothetical protein SORBIDRAFT_05g003840 [Sorghum bicolor]	0.072	46%	XP_002449038.1	-8,5747	-8,6011	-6,1766

#### 4.3.6.2 Análise de agrupamentos de perfis de expressão diferencial para a cultivar sensível (RB855453)

No cluster 3, onde agruparam-se os genes induzidos na cultivar sensível em relação à cultivar tolerante (Tabela 11), encontramos um gene com anotação para os fatores de transcrição básicos *helix-loop-helix* (HLH / bHLH), caracterizados por duas alfa-hélices conectadas por um loop, os quais pertencem a uma família de reguladores de transcrição presentes em três reinos eucarióticos. Muitas funções diferentes foram identificadas para estas proteínas em animais, incluindo o controle da proliferação de células e desenvolvimento de linhagens celulares específicas (HEIM et al, 2003).

Em plantas, pouco se sabe sobre a família bHLH. Vantini et al. (2015) trabalhando com cana-de-açúcar e déficit hídrico, identificaram a expressão do fator de transcrição bHLH com três dias de estresse hídrico. Em outro experimento, Heim et al. (2003), determinaram que existem 133 genes bHLH em *Arabidopsis thaliana*, e que pelo menos 113 deles são expressos. Em seu estudo, doze subfamílias foram identificadas e supõe-se que possuem papéis diferentes na célula vegetal e desenvolvimento dos tecidos, bem como o metabolismo das plantas. Exemplos de membros da família bHLH de *A. thaliana* incluem as proteínas GLABRA3 (GL3), importantes na diferenciação celular dos tricomas, o “Phytochrome interacting factor 3” (PIF3), necessário para a sinalização da luz mediada pelo foto-receptor do fitocromo B (PhyB) e os produtos dos genes AtbHLH044/BEE1, AtbHLH058/BEE2 e AtbHLH050/BEE3 que têm sido relacionados a múltiplas vias de regulação do crescimento e desenvolvimento das plantas.

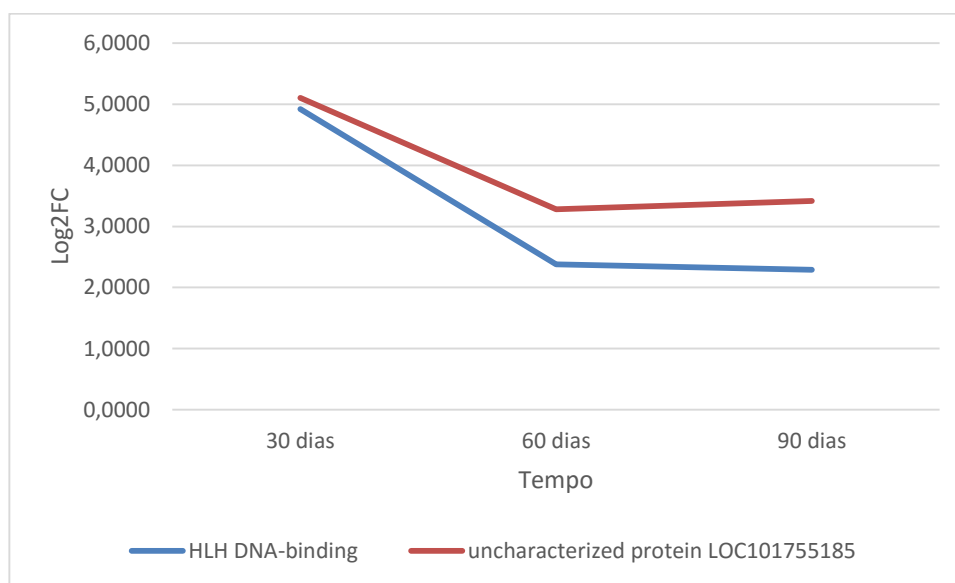
Hao et al (2012), também trabalhando com *A. thaliana* concluíram que as bHLH junto com fatores que interagem com fitocromo regulam o alongamento celular e o desenvolvimento das plantas em resposta à luz e hormônios.

Foi encontrado também na cultivar sensível, uma proteína não caracterizada que apresenta uma região das proteínas fucosiltransferases, semelhante à família das glicosiltransferases, relacionados com a biossíntese de polissacarídeos da parede celular e ligação de porções de açúcar a várias moléculas pequenas, tais como hormônios e flavonóides (KEEGSTRA; RAIKHEL, 2001).

Como todas as glicosiltransferases, as fucosiltransferases (FUTs) precisam de dois substratos para catalisar a reação: um doador de açúcar e um receptor (MARTINEZ-DUNCKE et al., 2003).

As descrições das capacidades dessas proteínas em organismos-modelo como *Arabidopsis*, *Caenorhabditis* e *Drosophila* têm sido escassos, e sua especificidade ainda permanece desconhecida. Acredita-se que a biossíntese de celulose (principal componente da parede celular de plantas, responsável pela rigidez e firmeza das plantas) e embriogênese normal em plantas exigem certas enzimas que são partes da via de glicosilação (WILSON, 2002).

A expressão desses genes ao longo do tempo pode ser observada no gráfico da Figura 23. Podemos verificar que a expressão deles é alta aos 30 dias, porém, começa a decair até os 60 e se mantém constante até 90 dias.



**Figura 23.** Comportamento de dois genes induzidos na cultivar sensível em relação à cultivar tolerante (Cluster 3) ao longo do tempo (30, 60 e 90 dias).

**Tabela 10.** Descrição dos genes encontrados no Cluster 3 após blastx contra banco de dados não redundante (“nr”)

CLUSTER 3 – Cultivar Sensível								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g81099	1147	Nenhuma similaridade significativa identificada	-	-	-	5,9189	3,3026	2,5700
c0_g64496	888	hypothetical protein SORBIDRAFT_0015s007030 [Sorghum bicolor]	2,00E-78	44%	XP_002489165.1	5,9157	2,9069	2,4875
c0_g101696	1357	Nenhuma similaridade significativa identificada	-	-	-	5,8165	3,3983	2,3922
c0_g131897	701	Nenhuma similaridade significativa identificada	-	-	-	5,6661	2,8707	2,1557
c0_g126663	819	PREDICTED: probable thiol methyltransferase 2-like isoform X2 [Setaria italica]	3,00E-04	48%	XP_004964819.1	5,5179	2,8406	2,6684
c0_g4620	1944	hypothetical protein SORBIDRAFT_01g023090 [Sorghum bicolor]	0.0	62%	XP_002464669.1	5,4848	3,9771	2,2382
c0_g81096	270	hypothetical protein SORBIDRAFT_09g007200 [Sorghum bicolor]	5,00E-47	96%	XP_002439466.1	5,3888	3,8201	2,6793
c0_g102265	475	Nenhuma similaridade significativa identificada	-	-	-	5,2593	3,8867	2,9011
c0_g13692	1680	hypothetical protein ZEAMMB73_767827 [Zea mays]	2,00E-47	93%	AFW82827.1	5,2191	3,3007	3,2470
c0_g44894	250	hypothetical protein [Labrenzia sp. DG1229]	1.7	30%	WP_029065296.1	5,2028	3,2480	2,9755
c0_g86957	1126	retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	6,00E-05	56%	AAY23271.1	5,1829	3,9685	2,5621
c0_g74864	402	hypothetical protein [Campylobacter gracilis]	2.3	34%	WP_005871310.1	5,1627	2,8223	2,2549
c0_g122034	433	PREDICTED: beta-glucosidase 32-like [Setaria italica]	1,00E-04	100%	XP_004957322.1	5,1257	4,4213	2,3162
c0_g101697	407	Nenhuma similaridade significativa identificada	-	-	-	5,0076	3,5611	3,0264
c0_g33179	350	tetracycline resistance protein TetA [Jeotgalibacillus campisalis]	3.9	36%	WP_041053605.1	4,9840	3,8027	3,1262
c0_g63918	1065	PREDICTED: uncharacterized protein LOC100194277 isoform X2 [Zea mays]	0.011	92%	XP_008650957.1	4,9833	2,5625	2,1417
c0_g52277	1109	Nenhuma similaridade significativa identificada	-	-	-	4,8506	3,2851	3,2318

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g109655	520	hypothetical protein SORBIDRAFT_08g015425 [Sorghum bicolor]	5,00E-46	85%	XP_002442168.1	4,9674	3,9772	3,1233
c0_g22928	813	putative HLH DNA-binding domain superfamily protein [Zea mays]	6,00E-38	82%	AFW60293.1	4,9217	2,3810	2,2911
c0_g79550	351	hypothetical protein SORBIDRAFT_08g000535 [Sorghum bicolor]	5,00E-13	89%	XP_002442654.1	4,8562	3,8163	3,1346
c0_g39758	389	retrotransposon protein, putative, Ty1-copia subclass [Oryza sativa Japonica Group]	1,00E-35	61%	ABA97865.1	4,5191	4,7603	2,7825
c0_g72454	778	PREDICTED: zinc finger BED domain-containing protein RICESLEEPER 1-like [Elaeis guineensis]	3,00E-14	40%	XP_010924127.1	4,2695	4,3393	2,8649
c0_g124586	905	uncharacterized protein LOC100273715 [Zea mays]	0.010	79%	NP_001141598.1	4,2505	4,8299	2,9414
c0_g14298	227	hypothetical protein SCLCIDRAFT_1209376 [Scleroderma citrinum Fouq A]	0.10	36%	KIM67999.1	4,2146	4,5065	2,7533
c0_g108388	3065	Disease resistance protein RPM1 [Aegilops tauschii]	0.0	57%	EMT12200.1	4,2029	5,0410	2,8484
c0_g6517	246	universal stress protein [Herbaspirillum sp. TSA66]	6.4	33%	WP_040040294.1	4,1207	4,5906	2,7926
c0_g28744	1137	Nenhuma similaridade significativa identificada	-	-	-	4,1156	4,3075	3,0909
c0_g88729	329	Nenhuma similaridade significativa identificada	-	-	-	4,0899	4,3883	3,1016
c0_g18393	642	hypothetical protein SORBIDRAFT_0012s017630 [Sorghum bicolor]	6,00E-59	85%	XP_002489223.1	4,0880	4,3991	3,5813
c0_g66682	284	Nenhuma similaridade significativa identificada	-	-	-	4,0823	4,7752	2,6280
c0_g130547	1426	Nenhuma similaridade significativa identificada	-	-	-	4,0758	4,5150	3,2446
c0_g9885	4089	ypothetical protein SORBIDRAFT_04g014351 [Sorghum bicolor]	0.0	60%	XP_002453763.1	4,0590	4,4009	3,5719
c0_g50067	1221	Nenhuma similaridade significativa identificada	-	-	-	4,0572	4,1144	3,7850
c0_g111579	1211	Nenhuma similaridade significativa identificada	-	-	-	4,0413	4,1623	2,8251

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g102774	1065	hypothetical protein SORBIDRAFT_02g005880 [Sorghum bicolor]	2,00E-06	73%	XP_002461647.1	4,0393	4,6345	3,4188
c0_g44856	311	hypothetical protein [Vibrio litoralis]	1.4	31%	WP_027695645.1	4,0202	4,9914	2,9313
c0_g118078	3553	Os08g0205150 [Oryza sativa Japonica Group]	0.0	50%	NP_001175434.1	3,9893	3,9921	3,6998
c0_g25084	1167	hypothetical protein SORBIDRAFT_07g028200 [Sorghum bicolor]	2,00E-90	61%	XP_002444798.1	3,9721	4,2315	2,4183
c0_g105607	590	Nenhuma similaridade significativa identificada	-	-	-	3,9328	4,0845	3,5591
c0_g66309	262	chemotaxis protein [Rhizobium nepotum]	6.9	34%	WP_045019663.1	3,8985	3,9148	3,6300
c0_g914	1209	Nenhuma similaridade significativa identificada	-	-	-	3,8558	3,8574	4,0660
c0_g56716	373	retrotransposon protein, putative, Ty1-copia subclass [Oryza sativa Japonica Group]	0.003	46%	ABG00000.1	3,6835	4,0213	4,0847
c0_g39320	777	Nenhuma similaridade significativa identificada	-	-	-	3,5508	3,8699	4,1147
c0_g27322	1221	uncharacterized protein LOC100191535 [Zea mays]	0.0	89%	NP_001130438.1	3,4959	3,5756	4,8271
c0_g123425	1434	PREDICTED: uncharacterized protein LOC103637593 [Zea mays]	1,00E-44	46%	XP_008657975.1	3,4912	4,0308	4,1058
c0_g94946	1161	PREDICTED: uncharacterized protein LOC100191246 isoform X2 [Zea mays]	4,00E-39	77%	XP_008680950.1	3,4754	3,7585	4,4910
c0_g56231	1610	hypothetical protein SORBIDRAFT_03g001370 [Sorghum bicolor]	0.0	89%	XP_002454918.1	3,4517	3,5797	3,9476
c0_g42979	2632	hypothetical protein SORBIDRAFT_07g006250 [Sorghum bicolor]	0.0	95%	XP_002445226.1	3,4488	3,9709	4,3102
c0_g15307	1611	hypothetical protein SORBIDRAFT_02g036440 [Sorghum bicolor]	1,00E-11	78%	XP_002463025.1	3,4139	3,6359	3,8084
c0_g92382	313	hypothetical protein SORBIDRAFT_0010s012040 [Sorghum bicolor]	2,00E-31	97%	XP_002489311.1	3,3790	4,0987	4,1789
c0_g38444	844	Nenhuma similaridade significativa identificada	-	-	-	3,3367	3,5851	3,9154
c0_g113797	366	Nenhuma similaridade significativa identificada	-	-	-	3,2078	3,2653	3,3192

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g62151	2048	hypothetical protein SORBIDRAFT_08g001950 [Sorghum bicolor]	0.0	87%	XP_002442734.1	3,3188	3,4717	3,6510
c0_g97325	477	hypothetical protein SORBIDRAFT_05g002050 [Sorghum bicolor]	4,00E-06	96%	XP_002450216.1	3,2243	3,8633	3,9572
c0_g95471	402	hypothetical protein SORBIDRAFT_01g014700 [Sorghum bicolor]	9,00E-14	73%	XP_002464235.1	3,1670	4,2038	4,3754
c0_g112818	459	Nenhuma similaridade significativa identificada	-	-	-	3,1564	3,4229	3,5399
c0_g3166	750	Nenhuma similaridade significativa identificada	-	-	-	3,1380	3,2303	3,5714
c0_g68606	441	hypothetical protein [Oryza sativa Japonica Group]	8,00E-05	83%	AAG60187.1	3,1239	3,7414	3,9592
c0_g18784	469	PPK-1 [Caenorhabditis elegans]	6.3	35%	NP_491576.2	3,1087	4,0614	4,3085
c0_g80016	1136	Nenhuma similaridade significativa identificada	-	-	-	3,0980	3,3497	3,3626
c0_g115272	331	hypothetical protein [Pantoea sp. MBLJ3]	3.8	38%	WP_039661838.1	7,0218	2,3203	2,4425
c0_g90525	327	Syntaxin-22 [Triticum urartu]	6.9	58%	EMS53062.1	6,8633	2,3507	2,5430
c0_g33315	1427	Nenhuma similaridade significativa identificada	-	-	-	6,2938	2,2268	2,6912
c0_g124723	269	Os02g0583300 [Oryza sativa Japonica Group]	5,00E-18	95%	NP_001173043.1	6,0541	2,6632	3,1197
c0_g31517	382	putative multidrug export ATP-binding/permease protein [Clostridium botulinum A1 str. CFSAN002368]	3.0	38%	EPS51386.1	5,8575	2,2581	3,2235
c0_g38357	674	Nenhuma similaridade significativa identificada	-	-	-	5,2472	2,6622	3,6771
c0_g124947	273	hypothetical protein SORBIDRAFT_08g004120 [Sorghum bicolor]	1,00E-46	94%	XP_002442868.1	5,2147	3,0515	3,5170
c0_g127328	1193	PREDICTED: uncharacterized protein LOC101755185 [Setaria italica]	4,00E-78	88%	XP_004954321.1	5,1053	3,2794	3,4186
c0_g47890	835	cyclase/dehydrase family protein [Saccharum hybrid cultivar R570]	0.008	100%	AGT15916.1	4,9776	3,2589	3,5483
c0_g43682	576	Nenhuma similaridade significativa identificada	-	-	-	4,8615	2,6940	2,7294



**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g9218	667	hypothetical protein SORBIDRAFT_04g009860 [Sorghum bicolor]	2,00E-72	80%	XP_002453651.1	4,8503	3,1635	3,5201
c0_g125792	1485	hypothetical protein [Corynebacterium imitans]	8.5	32%	WP_038594006.1	4,8066	3,4314	3,4555
c0_g68466	889	PREDICTED: uncharacterized protein LOC101778741 [Setaria italica]	2,00E-120	90%	XP_004977960.1	4,7745	3,0335	3,8344
c0_g59832	689	Nenhuma similaridade significativa identificada	-	-	-	4,7469	3,1033	4,0622
c0_g52475	1087	putative non-LTR retroelement reverse transcriptase [Sorghum bicolor]	0.0	90%	ABE77191.1	4,7343	2,3448	3,7241
c0_g91396	370	DNA recombination protein RecN [Bacteroides graminisolvens]	5.6	35%	WP_024996112.1	4,7164	2,3484	3,5277
c0_g131264	495	Nenhuma similaridade significativa identificada	-	-	-	4,6996	3,4774	3,8020
c0_g107465	218	hypothetical protein, variant [Pneumocystis murina B123]	4.6	39%	XP_007874266.1	4,6887	2,8832	3,4676
c0_g31321	1505	Nenhuma similaridade significativa identificada	-	-	-	4,6556	3,1661	3,3340
c0_g87132	882	Nenhuma similaridade significativa identificada	-	-	-	4,6409	3,5097	3,5210

No cluster 4, também da cultivar sensível, foi encontrada uma proteína não caracterizada (Tabela 12), que apresenta uma região *AAI\_LTSS: Alpha-Amylase Inhibitors (AAI), Lipid Transfer (LT) and Seed Storage (SS) Protein family*, uma família de proteínas única de plantas superiores que inclui inibidores de alfa-amilase, proteínas de transferência de lipídios, proteínas de armazenagem de sementes.

As proteínas inibidoras de  $\alpha$ -amilase (AAI) funcionam como inibidoras de enzimas digestivas de insetos-praga. Os principais inibidores de  $\alpha$ -amilase conhecidos são os encontrados no feijão (*Phaseolus vulgaris*), trigo (*Triticum aestivum*) e milho (*Zea mays*). Os mecanismos de interação e especificidade dos inibidores com as enzimas digestivas são complexos e ainda não foram completamente elucidados. O que se sabe é que uma vez que a enzima  $\alpha$ -amilase é inibida, a assimilação de nutrientes pelos insetos é reduzida e, conseqüentemente, o desenvolvimento dessas pragas é afetado (JÚNIOR; LAZZARI; JÚNIOR, 2006).

As proteínas de transferência de lipídios (LTP) são proteínas básicas presentes em quantidades elevadas em plantas superiores. Alguns papéis são atribuídos a essas proteínas, como: participação na formação de cutina, embriogênese, reações de defesa contra fitopatógenos, simbiose e adaptação das plantas a diferentes condições ambientais (KADER, 1996).

As proteínas de armazenamento em sementes (SS) por sua vez, são responsáveis pelo armazenamento de nutrientes. Em soja, as duas proteínas de armazenamento principais são a glicina e  $\beta$ -conglucina, representando cerca de 70-80% das proteínas totais. Estas são em grande parte responsáveis pelas propriedades nutricionais e físico-químicas da soja (NATARAJAN et al., 2006).

Juntas essas proteínas desempenham um papel importante ao defender as plantas de insetos e patógenos, transporte lipídico entre as membranas intracelulares e armazenamento de nutrientes, necessários para o crescimento e desenvolvimento das plantas sob condições de estresse.

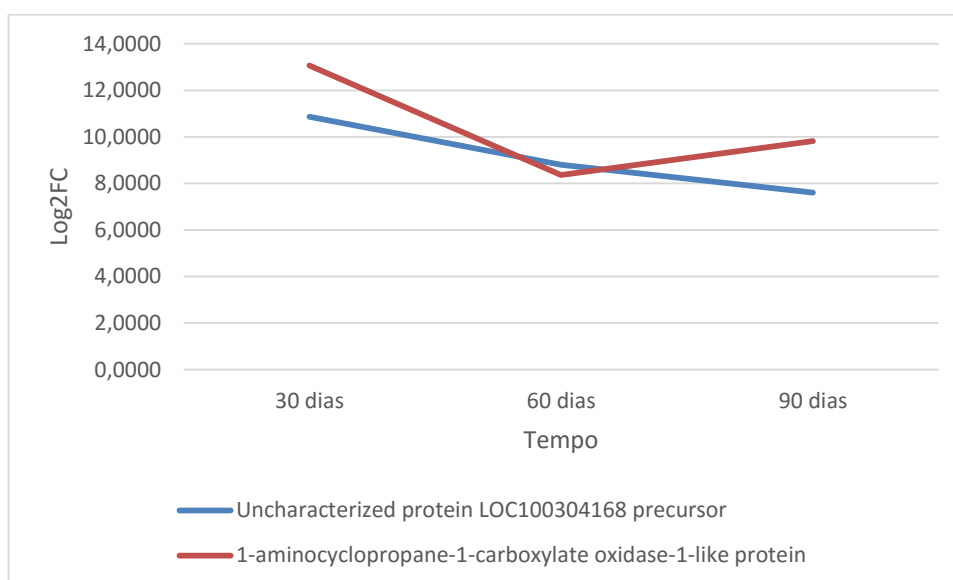
Também encontramos na cultivar sensível, uma enzima aminociclopropano carboxilato oxidase (ACO / ACC oxidase), responsável por catalisar o último passo na biossíntese de etileno, importante hormônio vegetal responsável pelo crescimento e desenvolvimento das plantas, tais como germinação, amadurecimento e senescência de frutas (MIRICA; KLINMAN, 2008).

Em tomate, as ACC oxidase se acumulam em vários tecidos, em diferentes fases de desenvolvimento e, durante a senescência das folhas, frutos e flores (BARRY et al., 1996).

Em 2014, em um estudo com milho, foi realizado o silenciamento dos genes ACC para modular os níveis de síntese de etileno. Os resultados demonstraram que a redução na via de biossíntese de etileno podem melhorar a produção de grãos, já que inibe a senescência dos mesmos (HABBEN et al., 2014),

Na cultivar sensível, esse gene provavelmente estaria atuando de forma a aumentar a síntese de etileno induzindo o amadurecimento e senescência de folhas.

O perfil de expressão dos três genes discutidos para o cluster 4 encontra-se no gráfico da Figura 24.



**Figura 24.** Comportamento de dois genes induzidos na cultivar sensível em relação à cultivar tolerante (Cluster 4) ao longo do tempo (30, 60 e 90 dias).

**Tabela 11.** Descrição dos genes encontrados no Cluster 4 após blastx contra banco de dados não redundante (“nr”)

CLUSTER 4 – Cultivar Sensível								
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g32784	441	Hypothetical protein OsJ_05629 [Oryza sativa Japonica Group]	0.056	78%	EEE56442.1	11,4477	7,4199	7,0024
c0_g66193	975	Ubiquitin carboxyl-terminal hydrolase 2 [Aegilops tauschii]	1.2	44%	EMT16654.1	11,1440	9,3202	6,7374
c0_g96169	2891	Hypothetical protein SORBIDRAFT_03g013172 [Sorghum bicolor]	7,00E-126	70%	XP_002455561.1	10,9887	8,7860	7,3622
c0_g128781	1457	PREDICTED: uncharacterized protein LOC101758390 [Setaria italica]	3,00E-64	38%	XP_004986728.1	10,9034	10,3711	8,7479
c0_g106602	2014	Uncharacterized protein LOC100304168 precursor [Zea mays]	1,00E-39	77%	NP_001159098.1	10,8684	8,7954	7,6029
c0_g104721	1282	Hypothetical protein SORBIDRAFT_05g024910 [Sorghum bicolor]	7,00E-118	74%	XP_002451139.1	10,4844	8,7212	5,9333
c0_g84454	216	PREDICTED: protein FAM179B [Solanum lycopersicum]	1.6	55%	XP_004229496.1	10,4120	8,8443	7,5741
c0_g51225	2624	Hypothetical protein SORBIDRAFT_06g014936 [Sorghum bicolor]	3,00E-156	55%	XP_002446373.1	10,3945	8,5836	7,8075
c0_g128112	257	Nenhuma similaridade significativa encontrada	-	-	-	10,3836	6,8103	4,9181
c0_g16825	3371	Disease resistance protein RPM1 [Triticum urartu]	5,00E-108	35%	EMS48852.1	10,3381	9,2284	8,8310
c0_g119023	373	Nenhuma similaridade significativa encontrada	-	-	-	10,3338	5,9744	5,6809
c0_g117219	1807	Nenhuma similaridade significativa encontrada	-	-	-	10,3316	8,5370	7,9759
c0_g125349	713	Hypothetical protein [Nocardia asiatica]	2.7	34%	WP_043724387.1	10,2853	8,7576	6,8235
c0_g131104	1236	Hypothetical protein SORBIDRAFT_02g026873 [Sorghum bicolor]	2,00E-04	83%	XP_002462504.1	10,1212	8,5068	6,6916
c0_g10900	824	Hypothetical protein TRIUR3_12404 [Triticum urartu]	1,00E-05	57%		9,9890	9,0474	8,1546
c0_g792	361	(Fe-S)-binding protein [Desulfomicrobium baculatum]	1.7	34%	WP_015773911.1	9,8654	9,1542	7,0504
c0_g64467	3850	Putative gag-pol polyprotein [Zea mays]	0.0	58%	AAL76001.1	9,8488	9,0430	8,3937

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g38772	484	Nenhuma similaridade significativa encontrada	-	-	-	9,8219	5,7737	3,9214
c0_g80983	229	Nenhuma similaridade significativa encontrada	-	-	-	9,8092	9,6999	7,9629
c0_g116630	1731	Uncharacterized protein LOC100303793 [Zea mays]	1,00E-04	37%	NP_001158899.1	9,7837	9,0541	8,4490
c0_g95689	1258	1-aminocyclopropane-1-carboxylate oxidase-1-like protein [Aegilops tauschii]	3,00E-136	79%	EMT14845.1	13,0678	8,3571	9,8153
c0_g50966	952	TPA: hypothetical protein ZEAMMB73_810828 [Zea mays]	5,00E-07	56%	DAA44593.1	12,0466	8,7753	9,8499
c0_g4828	372	Nenhuma similaridade significativa encontrada	-	-	-	11,8095	4,9504	5,9442
c0_g5402	967	Putative LRR receptor-like serine/threonine-protein kinase [Aegilops tauschii]	6,00E-71	67%	EEE51729.1	11,6572	6,4857	6,8917
c0_g108511	844	Hypothetical protein SORBIDRAFT_04g020775 [Sorghum bicolor]	2,00E-58	94%	XP_002452158.1	11,5795	6,4793	8,3098
c0_g4827	1088	TPA: hypothetical protein ZEAMMB73_921959 [Zea mays]	8,00E-08	83%	DAA39512.1	11,5379	8,6997	9,1152
c0_g125033	2222	Hypothetical protein SORBIDRAFT_10g021935 [Sorghum bicolor]	2,00E-43	59%	XP_002437143.1	11,3475	7,7569	8,1260
c0_111604	1251	Nenhuma similaridade significativa encontrada	-	-	-	10,5739	7,9142	8,8924
c0_g102552	827	PREDICTED: kinesin-like protein KIF15 isoform X2 [Poecilia reticulata]	6.9	33%	XP_008398271.1	10,5372	8,4188	8,5030
c0_g16315	2027	PREDICTED: disease resistance RPP13-like protein 4 [Zea mays]	0.0	77%	XP_008678367.1	10,4068	7,0419	7,9352
c0_g29568	474	Nenhuma similaridade significativa encontrada	-	-	-	10,0704	10,5510	7,1977
c0_g19919	3284	Receptor-like protein kinase 5 [Zea mays]	0.0	48%	XP_008659674.1	10,0395	8,7200	10,0497
c0_g76354	1841	Hypothetical protein SORBIDRAFT_05g000945 [Sorghum bicolor]	4,00E-96	53%	XP_002448886.1	10,3438	7,7477	8,8148
c0_g113469	1881	PREDICTED: uncharacterized protein LOC102719535 [Oryza brachyantha]	7,00E-160	52%	XP_006663597.1	10,2825	8,2029	9,4218
c0_g54673	2147	Cysteine-rich receptor-like protein kinase 10.	6,00E-111	36%	EMS62841.1	10,1953	9,7442	10,0569
c0_g54290	204	Nenhuma similaridade significativa encontrada	-	-	-	9,9517	9,6132	9,8765

Continuação									
Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC			
						30 dias	60 dias	90 dias	
c0_g56715	796	Hypothetical protein SORBIDRAFT_08g018875 [Sorghum bicolor]	4,00E-66	71%	XP_002442365.1	10,1176	9,4396	9,6290	
c0_g103153	1219	Cysteine-rich receptor-like protein kinase 27 [Aegilops tauschii]	1,00E-46	47%	EMT16994.1	10,0189	8,7830	9,7512	
c0_g111976	1019	Hypothetical protein SORBIDRAFT_05g022730 [Sorghum bicolor]	1,00E-172	92%	XP_002451019.1	9,9947	6,7139	7,0383	
c0_g85218	1932	Nenhuma similaridade significativa encontrada	-	-	-	9,9076	9,9944	8,3383	
c0_g29568	474	Nenhuma similaridade significativa encontrada	-	-	-	10,0704	10,5510	7,1977	
c0_g85218	1936	Nenhuma similaridade significativa encontrada	-	-	-	9,9076	9,9944	8,3383	
c0_g106877	1566	hypothetical protein SORBIDRAFT_05g013630 [Sorghum bicolor]	2,00E-39	38%	EEC76001.1	9,8093	9,9275	7,6269	
c0_g119038	1552	PREDICTED: putative nuclease HARBI1-like [Setaria italica]	3,00E-87	66%	XP_004961062.1	9,5245	9,5790	7,9283	
c0_g42699	1415	UDP-glucose pyrophosphorylase [Saccharum officinarum]	1.1	54%	ACM45950.1	9,5062	9,9041	8,0886	
c0_g28677	3253	PREDICTED: putative disease resistance protein RGA3 [Brachypodium distachyon]	0.0	62%	XP_003580978.1	9,4771	9,5189	8,7468	
c0_g73045	525	PREDICTED: uncharacterized protein LOC101767209 [Setaria italica]	6,00E-06	35%	AFW77842.1	9,4703	9,5950	8,9674	
c0_g129730	4074	PREDICTED: putative disease resistance protein RGA3 [Brachypodium distachyon]	0.0	52%	XP_003580978.1	9,2241	9,6348	8,4085	
c0_g125814	4498	retrotransposon protein, putative, unclassified [Oryza sativa Japonica Group]	0.0	55%	AAX95863.1	9,2224	9,6696	8,0857	
c0_g97046	2531	putative NBS-LRR disease resistance protein homologue [Oryza sativa Japonica Group]	3,00E-132	39%	BAD03231.1	9,2129	9,9466	9,5113	
c0_g130582	1040	hypothetical protein OsI_09551 [Oryza sativa Indica Group]	2,00E-07	53%	EEC74298.1	9,2126	9,6616	7,6102	
c0_g56258	3975	putative gag-pol polyprotein [Zea mays]	0.0	53%	AAL76001.1	9,1890	9,6235	9,9215	
c0_g105882	1395	Cysteine-rich receptor-like protein kinase 26 [Triticum urartu]	1,00E-112	55%	EMS61842.1	9,1591	9,5242	8,1388	
c0_g73335	412	Nenhuma similaridade significativa encontrada	-	-	-	9,1561	9,6772	8,5492	

Continuação

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g11847	459	hypothetical protein SORBIDRAFT_04g003490 [Sorghum bicolor]	7.7	50%	XP_002451541.1	9,0320	11,0976	10,2039
c0_g10436	404	Nenhuma similaridade significativa encontrada	-	-	-	9,0144	9,0281	8,2961
c0_g72506	1529	Nenhuma similaridade significativa encontrada	-	-	-	8,9490	9,2975	7,5491
c0_g128571	658	hypothetical protein SORBIDRAFT_01g024620 [Sorghum bicolor]	3,00E-08	80%	XP_002467312.1	8,8787	9,3534	8,5292
c0_g49982	909	hypothetical protein SORBIDRAFT_08g018875 [Sorghum bicolor]	8,00E-112	72%	XP_002442365.1	8,8776	9,2583	7,3515
c0_g128686	1419	hypothetical protein SORBIDRAFT_10g024140 [Sorghum bicolor]	1,00E-08	41%	XP_002461558.1	8,8650	9,2504	7,5238
c0_g56258	3975	PREDICTED: uncharacterized protein LOC100382585 isoform X1 [Zea mays]	0.0	53%	XP_008655231.1	9,1890	9,6235	9,9215
c0_g122010	582	PREDICTED: uncharacterized protein LOC101760134 [Setaria italica]	3,00E-56	92%	XP_004966023.1	8,8513	8,8982	9,6224
c0_g105296	629	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At4g37250 [Zea mays]	7,00E-13	51%	XP_008659794.1	8,6639	9,0914	9,2599
c0_g2805	2237	PREDICTED: uncharacterized protein LOC100823228 [Brachypodium distachyon]	0.0	74%	XP_010228395.1	8,4762	8,5050	9,1353
c0_g73417	534	Os07g0192700 [Oryza sativa Japonica Group]	7,00E-48	48%	NP_001059101.1	8,2041	8,9710	10,2132
c0_g80913	1812	Nenhuma similaridade significativa encontrada	-	-	-	8,0893	8,5171	8,6168
c0_g41975	479	PREDICTED: LOW QUALITY PROTEIN: anoctamin-8 [Columba livia]	0.72	28%	XP_005505577.1	7,8790	8,0359	8,2324
c0_g17215	1689	Nenhuma similaridade significativa encontrada	-	-	-	7,8380	7,8498	9,4097
c0_g72748	446	Nenhuma similaridade significativa encontrada	-	-	-	7,7219	8,0726	9,3584
c0_g11033	1574	PREDICTED: death-associated protein kinase 3-like [Setaria italica]	0.0	67%	XP_004979753.1	7,6230	8,3662	8,7767
c0_g53164	522	hypothetical protein SORBIDRAFT_10g022750 [Sorghum bicolor]	2,00E-71	96%	XP_002437199.1	7,6074	7,9123	8,1942
c0_g67816	2166	hypothetical protein [uncultured Mediterranean phage uvMED]	0.041	50%	BAQ93888.1	7,6005	7,6149	7,8933

**Continuação**

Gene	Tamanho (bases)	Anotação	e-value	Identidade	Acesso	Log2FC		
						30 dias	60 dias	90 dias
c0_g133523	3297	putative leucine rich repeat containing protein kinase [Oryza sativa Japonica Group]	0.0	67%	AAL82659.1	7,5832	7,9903	8,1102
c0_g68437	407	Nenhuma similaridade significativa encontrada	-	-	-	7,5711	8,0192	8,1877
c0_g35433	1715	hypothetical protein OsJ_29490 [Oryza sativa Japonica Group]	2.4	21%	EEE69774.1	7,2336	8,4006	9,4100
c0_g1094	243	Nenhuma similaridade significativa encontrada	-	-	-	7,2252	7,5529	7,6870
c0_g133741	269	Nenhuma similaridade significativa encontrada	-	-	-	7,1613	7,3225	8,3450
c0_g71172	532	short-chain dehydrogenase [Mycobacterium asiaticum]	3.1	35%	WP_036360960.1	7,1152	7,2055	7,2881
c0_g85769	216	hypothetical protein SORBIDRAFT_05g024306 [Sorghum bicolor]	0.22	57%	XP_002449844.1	7,1037	7,9291	7,9842
c0_g79776	836	hypothetical protein PFICI_02503 [Pestalotiopsis fici W106-1]	0.71	30%	XP_007829275.1	7,0862	7,6059	8,8521



## 5 CONCLUSÃO

Com base nos resultados obtidos neste trabalho, pode-se concluir que:

A técnica de RNA-Seq possibilitou a identificação de um grande número de transcritos em folhas de cana-de-açúcar, permitindo assim, avaliar os perfis de expressão gênica em resposta ao déficit hídrico prolongado nas duas cultivares analisadas.

O perfil de expressão gênica diferencial analisado neste trabalho mostrou que a cultivar sensível respondeu à seca ativando maior número de genes nos períodos iniciais (30 e 60 dias de estresse).

Por outro lado, a cultivar tolerante conseguiu reagir à seca sem a necessidade de ativar muitos genes. Esta ativação foi mais intensa aos 90 dias de estresse quando comparado à sensível, o que pode ser um indício de que a regulação do conteúdo gênico é mais importante na resposta ao déficit hídrico do que o próprio conteúdo gênico.

A validação via qRT-PCR de alguns genes diferenciamamente expressos e novas análises deverão ser realizadas a fim de comprovar e verificar, respectivamente, os genes envolvidos nos diferentes tempos e níveis de estresse (controle, moderado e severo), complementando ainda mais a compreensão das respostas das cultivares sob déficit hídrico prolongado.

## 6 REFERÊNCIAS

ADAMS, J. Transcriptome: connecting the genome to gene function. **Nature Education**, v. 1, p. 1, 2008.

ALEXANDERSSON, E.; FRAYSSE, L.; SJOVALL-LARSEN, S.; GUSTAVSSON, S.; FELLERT, M.; KARLSSON, M.; JOHANSON, U.; KJELLBOM, P. Whole gene family expression and drought stress regulation of aquaporins. **Plant Molecular Biology**, v. 59, p. 469–484, 2005.

AMBROSONE, A.; DI GIACOMO, M.; LEONE, A.; GRILLO, M. S.; COSTA, A. Identification of early induced genes upon water deficit in potato cell cultures by cDNA-AFLP. **Journal Plant Research - Regular paper**, v. 126, p. 169 - 178, 2013.

ARAP, M. A. **Estudo da proteína de choque térmico GRP78 para o desenvolvimento de um sistema de receptor-ligante para câncer de próstata**. 2003, 106f. Dissertação (Doutorado em Ciências) – Universidade de São Paulo, São Paulo, 2003.

ARAÚJO, E.L.; SIMABUKURO, E.A.; NOGUEIRA, R.J.M.C.; LAGÔA, A.M.M.A.; MELO, A.M.Y (Org.) (2005). **Estresses Ambientais: danos e/ou benefícios para as plantas**. 3ª edição. Recife, Sociedade Brasileira de Fisiologia Vegetal, v.1.

ASHBURNER, M.; BALL, C. A.; BLAKE, J. A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J. M.; DAVIS, A. P.; DOLINSKI, K.; DWIGHT, S. S.; EPPIG, J. T.; HARRIS, M. A.; HILL, D. P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J. C.; RICHARDSON, J. E.; RINGWALD, M.; RUBIN, G. M.; SHERLOCK, G. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n.1, p. 25–29, 2000.

AURY, J.-M.; CRUAUD, C.; BARBE, V.; ROGIER, O.; MANGENOT, S.; SAMSON, G.; POULAIN, J.; ANTHOUARD, V.; SCARPELLI, C.; ARTIGUENAVE, F.; WINCKER, P. High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. **BMC Genomics**, v. 9, n. 1, p. 603, 2008.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. **Nucleic Acids Research**, v. 24, p. 21-25, 1996.

BARBOSA R. R. **Estudo do efeito da inoculação da bactéria gluconacetobacter diazotrophicus na resposta de plantas de cana-de-açúcar ao estresse salino**. 2007. 55p. Monografia (Bacharelado em Biociências e Biotecnologia) – Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Rio de Janeiro, 2007.

BARLING, A.; SWAMINATHAN, K.; MITROS, T.; JAMES, B. T.; MORRIS, J.; NGAMBOMA, O.; HALL, M. C.; KIRKPATRICK, J.; ALABADY, M.; SPENCE, A. K.; HUDSON, M. E.; ROKHSAR, D. S.; MOOSE, S. P. A detailed gene expression study of the *Miscanthus* genus reveals changes in the transcriptome associated with the rejuvenation of spring rhizomes. **BMC Genomics**, v. 14, p. 864, 2013.

BARRETO, M. C. S.; FIGUEIREDO, M. V. B.; BURITY, H. A.; SILVA, M. L. R. B.; LIMA-FILHO, J. L. Produção e comportamento reológico de biopolímeros produzidos por rizóbios e caracterização genética. **Revista Brasileira Agrociência**, v. 17, n. 2-4, p. 221-227, 2011.

BARRY, C. S.; BLUME, B.; BOUZAYEN, M.; COOPER, W.; HAMILTON, A. J.; GRIERSON, D. Differential expression of the 1-aminocyclopropane-1-carboxylate oxidase gene family of tomato. **The Plant Journal**, v. 9, n. 4, p. 525 – 535, 1996.

BHATNAGAR-MATHUR, P.; VADEZ, V.; SHARMA, K. K. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. **Plant Cell Reports**, v. 27, p. 411-424, 2008.

BISELLI, J. M.; **Expressão diferencial de microRNAs em células mononucleares do sangue periférico de crianças com síndrome de Down**. 2011, 86p. Tese (Doutorado em Ciências da Saúde; Eixo Temático: Medicina e Ciências Correlatas)-Faculdade de Medicina de São José do Rio Preto (FAMERP), São José do Rio Preto, 2011.

BOCCA, S. N.; MAGIOLI, C.; MANGEON, A.; JUNQUEIRA, R. M.; CARDEAL, V.; MARGIS, R.; SACHETTO-MARTINS, G. Survey of glycine-rich proteins (GRPs) in the *Eucalyptus* expressed sequence tag database (ForEST). **Genetics and Molecular Biology**, v. 28, p. 608-624, 2005.

BOCHMAN, M. L.; SABOURIA, N.; ZAKIANA, V. A. Unwinding the functions of the Pif1 family helicases. **DNA Repair**, v. 9, n.3, p. 237–249, 2010.

BOGUSKI, M. S.; LOWE, T. M.; TOLSTOSHEV, C. M. dbEST--database for "expressed sequence tags". **Nature Genetics**, v. 4, n. 4, p. 332-333, 1993.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114 – 2120, 2014.

BORGES, J. C.; PEROTO, M. C.; RAMOS, C. H. I. Molecular chaperone genes in the sugarcane expressed sequence database (SUCEST). **Genetics and Molecular Biology**, v.24, p. 85-92, 2001.

BOULÉ, J-B.; ZAKIAN, V. A.; Roles of Pif1-like helicases in the maintenance of genomic stability. **Nucleic Acids Research**, v. 34, n. 15, p. 4147–4153, 2006.

BOYD, C.; HUGHES, T.; PYPAERT, M.; NOVICK, P. Vesicles carry most exocyst subunits to exocytic sites marked by the remaining two subunits, Sec3p and Exo70p. **Journal Cell Biology**, v. 167, p. 889–901, 2004.

BRAY, E. A. Plant responses to water deficit. **Trends in Plant Science**, v. 2, p. 48-54, 1997.

BUCHANAN, B.; GRUISSEM, W.; JONES, R. (eds). **Biochemistry & Molecular Biology of Plants**. American Society of Plant Biologists, Rockville, Maryland, p. 1159-1162, 2000.

CAGLIARI, T. C. **Análise da expressão de chaperonas moleculares em plantas e clonagem, purificação e caracterização inicial da proteínas Hsp100 e Hsp90 de cana-de-açúcar**. 2009, 216f. Dissertação (Doutorado em Biologia Funcional e Molecular) – Universidade Estadual de Campinas, Campinas, 2009.

CAMACHO, C.; COULOURIS, G.; AVAGYAN, V.; MA, N.; PAPADOPOULOS, J.; BEALER, K.; MADDEN, T. L. BLAST+: architecture and applications. **BMC Bioinformatics**, v. 10, p.421, 2009.

CAMBRAIA, J. Aspectos bioquímicos, celulares e fisiológicos dos estresses nutricionais em plantas. In: NOGUEIRA, R.J.M.C.; ARAÚJO, E. DE L.; WILLADINO, L.G.; CAVALCANTE, U.M.T. (eds.). **Estresses ambientais: danos e benefícios em plantas**. Recife: UFRPE, 2005. cap. 2 , p. 95-104.

CARDOSO-SILVA, C. B.; COSTA, E. A.; MANCINI, M. C.; BALSALOBRE, T. W.; CANESIN, L. E. C.; PINTO, L. R.; CARNEIRO, M. S.; GARCIA, A. A. F.; SOUZA, A. P.; VICENTINI, R. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. **PloS One**, v. 9, n. 2, 2014.

CARLIN, S.D.; SANTOS, D.M.M. Indicadores fisiológicos da interação entre déficit hídrico e acidez do solo em cana-de-açúcar. **Pesquisa Agropecuária Brasileira**, v.44, n.9, p.1106-1113, 2009.

CARVALHO, M. H. C. Drought and reactive oxygen species. Productiton, scavenging and signaling. **Plant Signalling & Behavior**, v. 3, p. 156 – 165, 2008.

CARVALHO, M. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas, **Ciência Rural**, Santa Maria, v. 40, n.3, p.735-744, 2010.

CASAGRANDE, E. C.; FARIAS, J. R. B.; NEUMAIER, N.; OYA, T.; PEDROSO, J.; MARTINS, P. K.; BRETON, M. C.; NEPOMUCENO, A. L. Expressão gênica diferencial durante déficit hídrico em soja. **Revista Brasileira de Fisiologia Vegetal**, v. 13, n. 2, p. 168-184, 2001.

CASTRO, P. H.; TAVARES, R. M.; BEJARANO, E. R.; AZEVEDO, H. SUMO, a heavyweight player in plant abiotic stress responses. **Cellular and Molecular Life Science**, v. 69, p. 3269–3283, 2012.

CASU, R. E.; JARMEY, J. M.; BONNETT, G. D.; MANNERS, J. M. Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. **Functional Integrated Genomics**, v. 7, p. 153 – 167, 2007.

CATALA, R.; OUYANG, J.; ABREU, I. A.; HU, Y.; SEO, H.;ZHANG, X.; CHUA, N-H. The Arabidopsis E3 SUMO Ligase SIZ1 Regulates Plant Growth and Drought Responses. **The Plant Cell**, v. 19, p. 2952–2966, 2007.

CHAUMONT, F.; TYERMAN, S. D. Aquaporins: Highly Regulated Channels Controlling Plant Water Relations. **Plant Physiology**, v. 164, p. 1600–1618, 2014.

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

CHERIAN, S.; REDDY, M. P.; FERREIRA, R. B. Transgenic plants with improved dehydration-stress tolerance: progress and future prospects. *Biologia Plantarum*, v. 50, p. 481-495, 2006.

CHOWDHURY, E. K.; CHOI, B.; CHO, B-K.; KIM, J. K.; PARK, S. U.; NATARAJAN, S.; LIM, H. S.; BAE, H. Regulation of *4CL*, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions. **Plant Omics Journal**, v.6, n.4, p. 254 – 262, 2013.

CLEMENS, M. J.; BOMMER, U. A. Translational control: the cancer connection. **International Journal of Biochemistry & Cell Biology**, v. 31, p. 1-23, 1999.

COFRE, T.D.C.G. **Transcriptoma (RNA-seq) de laranja doce Valência (*Citrus sinensis* L. Osbeck) e Kumquat (*Fortunella* spp.) infectadas por *Xanthomonas citri* subsp. *Citri***. 2013. 148 f. Dissertação (Mestrado em Biociências e Biotecnologia) - Universidade Estadual Paulista – UNESP, Faculdade de Ciências Farmacêuticas de Araraquara, Araraquara, 2013.

COLLINS, F. S.; GREEN, E. D.; GUTTMACHER, A. E.; GUYER, M. S. A vision for the future of genomics research. **Nature**, v. 422, p. 835-847, 2003.

CONAB. Companhia Nacional de Abastecimento. Acompanhamento de safra brasileira – Cana-de-açúcar. 2014. Disponível em: <[http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14\\_12\\_19\\_09\\_02\\_49\\_boletim\\_cana\\_portugues\\_-\\_3o\\_lev\\_-\\_2014-15.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14_12_19_09_02_49_boletim_cana_portugues_-_3o_lev_-_2014-15.pdf)>. Acesso em 10 de março de 2015.

COSTA, V.; CASAMASSIMI, A.; CICCODICOLA, A. Nutritional genomics era: opportunities toward a genome-tailored nutritional regimen. **The Journal of Nutritional Biochemistry**, v. 21, n. 6, p. 457–467, 2010.

CRUZ, E. A. O. **Caracterização de transposases da família SchAT em cana-de-açúcar: estudo molecular e funcional**. 2012, 137f. Dissertação (Doutorado em Biotecnologia) – Universidade de São Paulo, Instituto de Ciências Biomédicas, São Paulo, 2012.

DANIELS, J.; ROACH, B.T. Taxonomy and evolution. In: HEINZ, D.J. **Sugarcane improvement through breeding**. New York: Elsevier, 1987. 84p.

DE SOUZA, A. P.; GASPAR, M.; DA SILVA, E. A.; ULIAN, E. C.; WACLAWOVSKY, A. J.; NISHIYAMA, M. Y.; DOS SANTOS, R. V.; TEIXEIRA, M. M.; SOUZA, G. M.; BUCKERIDGE, M. S. Elevated CO<sub>2</sub> increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane. **Plant Cell and Environment**, v. 31, p. 1116-1127, 2008.

DELGADO, A. A.; CÉSAR, M. A. A. **Elementos de tecnologia e engenharia do açúcar de cana**. Piracicaba: ESALQ, Depto. de Tecnologia Rural, p. 364, 1977.

DEPRÁ, M. **Estudo da mobilização de transposons através de transformação genética**. 2005, 78f. Dissertação (Mestrado em Genética e Biologia) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

DINIZ, M. C. **Associação de marcadores (AFLP) com caracteres de interesse agrônomo em clones elite de cana-de-açúcar (*Saccharum* spp)**. 2007. 104f. Dissertação (Mestrado em Genética na área de Genética) - Universidade Federal de Pernambuco, Recife, UFPE, Pernambuco, 2007.

DIXON, R. A.; PAIVA, N. L. Stress-induced phenylpropanoid metabolism. **The Plant Cell**, v. 7, p. 1085-1097, 1995.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. **Agroecologia da Cana-de-Açúcar**. 2011. Disponível em: <<http://www.cana.cnpm.embrapa.br/agroeco.html>>. Acesso em: 13 nov. 2014.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. **Entendendo a crise hídrica**. 2014. Disponível em : <<https://www.embrapa.br/aguanaagricultura/observatoriosafra20142015>> Acesso em: 12 março 2015.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. **O papel das mutações na compreensão da genética do milho**. 2010. Disponível em : <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/25545/1/Doc-100.pdf>> Acesso em: 20 junho 2015.

FAN, L.; LINKER, R.; GEPSTEIN, S.; TANIMOTO, E.; YAMAMOTO, R.; NEUMANN, P. M. Progressive inhibition by water deficit of cell wall extensibility and growth along the elongation zone of maize roots is related to increased lignin metabolism and progressive stelar accumulation of wall phenolics. **Plant Physiology**, v. 140, p. 603-612, 2006.

FEDATTO, L. M. **Caracterização de proteases extracelulares produzidas por *Xylella fastidiosa* de citros e videira**. 2004, 67f. Dissertação (Mestrado em Ecologia em agroecossistemas) – Universidade de São Paulo, Piracicaba, 2004.

FERREIRA, B. S. **Expressão diferencial de genes envolvidos na resposta ao estresse salino em cana-de-açúcar (*Saccharum* sp.) e arroz (*Oryza sativa*)**. 2008. 149 f. Dissertação (Doutorado em Biociências e Biotecnologia) - Universidade Estadual Do Norte Fluminense, Campos Dos Goytacazes, 2008.

FINGER, F.P.; HUGHES, T. E.; NOVICK, P. Sec3p is a spatial landmark for polarized secretion in budding yeast. **Cell**, v. 92, p. 559–571, 1998.

FRASER, C. M.; THOMPSON, M. G.; SHIRLEY, A. M.; RALPH, J.; SCHOENHERR, J. A.; SINLAPADECH, T.; HALL, M. C.; CHAPPLE, C. Related Arabidopsis Serine Carboxypeptidase-Like Sinapoylglucose Acyltransferases Display Distinct But Overlapping Substrate Specificities. **Plant Physiology**, v. 144, p. 1986–1999, 2007.

FU, X.; FU, N.; GUO, S.; YAN, Z.; XU, Y.; HU, H.; MENZEL, C.; CHEN, W.; LI, Y.; ZENG, R.; KHAITOVICH, P. Estimating accuracy of RNA-Seq and microarrays with proteomics. **BMC Genomics**. v. 10, p. 161, 2009. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-10-161>>.

GARG, R.; PATEL, R. K.; TVAGI, A. K.; JAIN, M. De Novo Assembly of Chickpea Transcriptome Using Short Reads for Gene Discovery and Marker Identification. **DNA Research**, v. 18, n. 1, p. 53-63, 2011.

GENTILE, A.; DIAS, L. I.; MATTOS, R. S.; FERREIRA, T. H.; MENOSSI, M. MicroRNAs and drought responses in sugarcane. **Plant Science**, v.6, p. 1- 13, 2015.

GENTLEMAN, R. C.; CAREY, V. J.; BATES, D. M.; BOLSTAD, B.; DETTLING, M.; DUDOIT, S.; ELLIS, B.; GAUTIER, L.; GE, Y.; GENTRY, J.; HORNIK, K.; HOTHORN, T.; HUBER, W.; IACUS, S.; IRIZARRY, R.; LEISCH, F.; LI, C.; MAECHLER, M.; ROSSINI, A. J.; SAWITZKI, G.; SMITH, C.; SMYTH, G.; TIERNEY, L.; YANG, J. Y.; ZHANG, J. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, n. 10, R80, 2004.

GOMES, B. M. M. F. **TreeHop: a method to improve orthology detection**. 2013. 60p. Dissertação (Mestrado em Bioinformática e Biologia Computacional) – Universidade de Lisboa, 2013.

GONÇALVES, E. R. **Fotossíntese, osmorregulação e crescimento inicial de quatro variedades de cana-de-açúcar submetida à deficiência hídrica**. 2008. 66 f. Dissertação (Mestrado em Agronomia) – Universidade Federal de Alagoas, Rio Largo, 2008.

GOODSTEIN, D. M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R. D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, p. 178-186, 2012.

GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACOEN, N.; GNIRKE, A.; RHIND, N.; DI PALMA, F.; BIRREN, B. W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature Biotechnology**, v. 29, n. 7, p. 644-654, 2011.

GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACOEN, N.; GNIRKE, A.; RHIND, N.; DI PALMA, F.; BIRREN, B. W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature Biotechnology**, v. 29, n. 7, p. 644–652, 2013.

GRATIVOL, C.; REGULSKI, M.; BERTALAN, M.; MCCOMBIE, W. R.; DA SILVA, F. R.; ZERLOTINI NETO, A.; VICENTINI, R.; FARINELLI, L.; HEMERLY, A. S.; MARTIENSSSEN, R. A.; FERREIRA, P. C. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. **Plant Journal**, v. 79, p. 162-172, 2014.

GUI, J.; SHEN, J.; LI, L. Functional Characterization of Evolutionarily Divergent 4-Coumarate: Coenzyme A Ligases in Rice. **Plant Physiology**, v. 157, p. 574–586, 2011.

GUPTA, V.; RAGHUVANSHI, S.; GUPTA, A.; SAINI, N.; GAUR, A.; KHAN, M. S.; GUPTA, R. S.; SINGH, J.; DUTTAMAJUMDER, S. K.; SRIVASTAVA, S.; SUMAN, A.; KHURANA, J. P.; KAPUR, R.; TYAGI, A. K. The water-deficit stress- and red-rot-related genes in sugarcane. **Funct Integr Genomics**, v. 10, p. 207-214, 2010.

HAAS B. J.; CHIN M.; NUSBAUM C.; BIRREN B. W.; LIVNY J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? **BMC Genomics**, v. 13, 2012.

HAAS, B. J.; PAPANICOLAOU, A.; YASSOURI, M.; GRABHERR, M.; BLOOD, P. D.; BOWDEN, J.; COUGER, M. B.; ECCLES, D.; LI, B.; LIEBER, M.; MACMANES, M. D.; OTT, M.; ORVIS, J.; POCHE, N.; STROZZI, F.; WEEKS, N.; WESTERMAN, R.; WILLIAM, T.; DEWEY, C. N.; HENSCHER, R.; LEDUC, R. D.; FRIEDMAN, N.; REGEV, A. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. **Nature Protocols**, v. 8, p. 1494–1512, 2013.



HABBen, J. E.; BAO, X.; BATE, N. J.; DEBRUIN, J. L.; DOLAN, D.; HASEGAWA, D.; HELENTJARIS, T. G.; LAFITTE, R. H.; LOVAN, N.; MO, H.; REIMANN, K.; SCHUSSLER, J. R. Transgenic alteration of ethylene biosynthesis increases grain yield in maize under field drought-stress conditions. **Plant Biotechnology Journal**, v. 12, p. 685–693, 2014.

HACHEZ, C.; ZELAZNY, E.; CHAUMONT, F. Modulating the expression of aquaporin genes in planta: A key to understand their physiological functions? **Biochimica et Biophysica Acta**, v.1758, p. 1142–1156, 2006.

HAMBERGER, B.; HAHLBROCK, K. The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapateactivating and three commonly occurring isoenzymes. **Proceedings of the National Academy of Sciences**, v. 101, n. 7, p. 2209–2214, 2004.

HAO, Y.; OH, E.; CHOI, G.; LIANG, Z.; WANG, Z-Y. Interactions between HLH and bHLH Factors Modulate Light-Regulated Plant Development. **Molecular Plant**, n. 5, v.3, p. 688 – 697, 2012.

HARRISON, C. GrpE, a nucleotide exchange Factor for DnaK. **Cell Stress & Chaperones**, v. 8, n. 3, p. 218–224, 2003.

HEIM, M. A.; JAKOBY, M.; WERBER, M.; MARTIN, C.; WEISSHAAR, B.; BAILEY, P. C. The Basic Helix–Loop–Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. **Molecular Biology and Evolution**, v.20, n. 5, p. 735–747, 2003.

HÖCK, J.; MEISTER, G. The Argonaute protein family. **Genome Biology**, v. 9, p. 210 – 218, 2008.

HOLMES, K. C. Structural biology: actin in a twist. **Nature**, v. 457, p. 389-390, 2009.

HOTTON, S. K.; CALLIS, J. Regulation of cullin RING ligases. **Annual Review Plant Biology**, v. 59, p. 467- 489, 2008.

INMAN-BAMBER, N. G.; SMITH, D. M. Water relations in sugarcane and response to water deficits. **Field Crops Research**, v. 92, p. 185-202, 2005.

ISKANDAR, H.M.; CASU, R.E.; FLETCHER, A.T.; SCHMIDT, S.; XU, J.; DONALD J MACLEAN, D.J.; MANNERS, J.M.; BONNETT, G.D. Identification of drought-response genes and a study of their expression during sucrose accumulation and water deficit in sugarcane culms. **BMC Plant Biology**, v. 11, n. 12, p. 1-14, 2011.

JAIN, M. Next generation sequencing technologies for gene expression profiling in plants. **Briefings in functional genomics**, v. 2, p. 63-70, 2011. Disponível em: <<http://dx.doi.org/10.1093/bfpg/elr038>>.

JALEEL, C. A.; MANIVANNAN, P.; SANKAR, B.; KISHOREKUMAR, A.; GOPI, R.; SOMASUNDARUM, R.; PANNEERSELVAN, R. Water deficit stress mitigation by calcium chloride in *Catharanthus roseus*: Effects on oxidative stress, praline metabolism and indole alkaloid accumulation. **Colloids and Surfaces B: Biointerfaces**, v.60, p.110-116, 2007.

JAMAUX, I.; STEINMETZ, A.; BELHASSEN, E. Looking for molecular and physiological markers of osmotic adjustment in sunflower. **New Phytologist**, v. 137, p. 117-127, 1997

JARZYNIAK, K. M.; JASINSKI, M. Membrane transporters and drought resistance—a complex issue. **Frontiers in Plant Science**, v. 5, p.687-701, 2014.

JIMÉNEZ, A.; HERNANDEZ, J. A.; DEL RIO, L. A.; SEVILHA, F. Evidence for the presence of the ascorbate-glutathione cycle in mitochondria and peroxisomes of pea leaves. **Plant Physiology**, v. 114, p. 275-284, 1997.

JOHNSON, E. S. Protein modification by SUMO. **Annual Review Biochemistry**, v. 73, p. 355 – 382, 2004.

JÚNIOR, A. L. M.; LAZZARI, S. M. N.; JÚNIOR, A. R. P. Inibidores de enzimas digestivas de insetos-praga. **Revista Acadêmica.**, Curitiba, v.4, n.1, p. 57-61, 2006.

KADER, J-C. Lipid-Transfer Proteins In Plants. **Plant Molecular Biology**, v. 47, p. 627–54, 1996.

KEEGSTRA, K.; RAIKHEL, N. Plant glycosyltransferases. **Plant Biology**, v. 4, p. 219–224, 2001.

KIDO, E. A.; FERREIRA NETO, J. R. C.; SILVA, R. L. O.; Pandolfi, V.; GUIMARÃES, A. C. R.; VEIGA, D. T.; CHABREGAS, S. M.; CROVELLA, S.; BENKO-ISEPPON, A. M. New insights in the sugarcane transcriptome responding to drought stress as revealed by supersage. **The Scientific World Journal**, v. 2012, p. 821062, 2012.

KIM, C.; LEE, T. H.; COMPTON, R. O.; ROBERTSON, J. S.; PIERCE, G. J.; PATERSON A. H. A genome-wide BAC end-sequence survey of sugarcane elucidates genome composition, and identifies BACs covering much of the euchromatin. **Plant Molecular Biology**, v. 81, p. 139-147, 2013.

KNOLL, A.; PUCHTA, H. The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. **Journal of Experimental Botany**, v. 62, n. 5, p. 1565–1579, 2011.

LANGMEAD, B.; SALZBERG, S. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357 – 359, 2012.

LEE, B. R.; KIM, K. Y.; JUNG, W. J.; AVICE, J. C.; OURRY, A.; KIM, T. H. Peroxidases and lignification in relation to the intensity of water-deficit stress in white clover (*Trifolium repens* L.). **Journal of Experimental Botany**, v. 58, p. 1271-1271, 2007.

LESTARI, P.; VAN, K.; KIM, M. Y.; HWANG, C. H.; LEE, B. W.; LEE, S.-H. Differentially expressed genes related to symbiotic association in a supernodulating soybean mutant and its wild-type. **Molecular Plant Pathology**, v. 7, n. 4, p. 235-247, 2006.

LI, S.; OS, G. M. A.; REN, S.; YU, D.; KETELAAR, T.; EMONS, A. M. C; LIU, C. Expression and Functional Analyses of EXO70 Genes in Arabidopsis Implicate Their Roles in Regulating Cell Type-Specific Exocytosis. **Plant Physiology**, v. 154, p. 1819–1830, 2010.

LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics Applications Note**, v. 22, n. 13, p. 1658–1659, 2006.

LIU, H.; WANG, X.; ZHANG, H.; YANG, Y.; GE, X.; SONG, F. A rice serine carboxypeptidase-like gene OsBISCPL1 is involved in regulation of defense responses against biotic and oxidative stress. **Gene**, v. 420, p. 57–65, 2008.

LIU, S.; CHEN, X.; HE, H.; ZHANG, X.; XIE, B.; YU, Y.; CHEN, B.; ZHOU, B.; ZHANG, Y. Structure and ecological roles of a novel exopolysaccharide from the arctic sea ice bacterium *pseudolateromonas* sp. Strain SM20310. **Appl. Environ. Microbiology**, v. 1. p. 224, 2013.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, p. 550, 2014.

LYRA, G. B.; BATISTA, E. L. S.; LYRA, G. B.; PEREIRA, C. R.; SILVA, L. D. B.; SILVA, G. M. **Coeficiente da cultura da cana-de-açúcar no estágio inicial de desenvolvimento em campos dos Goytacazes**, RJ. Irriga, Botucatu, v.17, n.1, p.102-113, 2012.

LYZENGA, W. J.; STONE, S. L. Abiotic stress tolerance mediated by protein ubiquitination. **Journal of Experimental Botany**, v. 63, n. 2, p. 599–616, 2012.

MACHADO, R.S.; RIBEIRO, R.V.; MARCHIORI, P.E.R.; MACHADO, D.F.S.P.; MACHADO, E.C.; LANDELL, M.G.A. Respostas biométricas e fisiológicas ao deficit hídrico em cana-de-açúcar em diferentes fases fenológicas. **Pesquisa Agropecuária Brasileira**, v.44, n.12, p.1575-1582, 2009.

MAGUIRE, B. A.; ZIMMERMANN, R. A. The ribosome in focus. **Cell**, v. 104, p. 813–816, 2001.

MAPA- Ministério da Agricultura. **Cana de açúcar. 2015**. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/cana-de-acucar>>. Acesso em: 09 julho 2015.

MARDIS, E. R. Next-generation DNA sequencing methods. **Annual Review of Genomics and Human Genetics**, vol. 9, p. 387–402, 2008.

MARGUERAT, S.; BAHLER, J. RNA-seq: from technology to biology. **Cellular and Molecular Life Sciences**, v. 67, p. 569–579, 2010. Disponível em: <<http://dx.doi.org/10.1007/s00018-009-0180-6>>.

MARQUES, M.V.; SILVA, A.M. **Genômica funcional: transcriptoma**. In: Luis Mir. (Org.). *Genômica*. São Paulo: Editora Atheneu, 2004.

MARTIN, J. Rnnotator: na automated *de novo* transcriptome assembly pipeline from stranded RNA-Seq reads. **BMC Genomics**, London, v. 11, n. 663, p. 1-8, 2010.

MARTINEZ-DUNCKER, I.; MOLLICONE, R.; CANDELIER, J-J.; BRETON, C.; ORIOL, R. A new superfamily of protein-O-fucosyltransferases,  $\alpha$ 2- fucosyltransferases, and  $\alpha$ 6-fucosyltransferases: phylogeny and identification of conserved peptide motifs. **Glycobiology**, v. 13, n. 12, p. 1C - 5C, 2003.

MARTINS, A. M. **Sequenciamento de DNA, montagem *de novo* do genoma e desenvolvimento de marcadores microssatélites, indels e SNPs para uso em análise genética de *Brachiaria ruziziensis***. 2013. 198 f. Dissertação (Doutorado em Biologia Molecular) – Universidade de Brasília , Instituto de Ciências Biológicas, Brasília, 2013.

MARTINS, N. G. S. **Os Fosfatos na Cana-de-Açúcar**. 2004. 87f. Dissertação (Mestrado em Ciências na área de Ciência e Tecnologia de Alimentos) – Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, Esalq/USP, São Paulo, 2004.

MCCARTHY, D. J.; CHEN, Y.; SMYTH, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. **Nucleic Acids Resource**, v. 40: 4288-4297, 2012.

MENESES, C. H. S. G.; LIMA, L. H. G. M.; LIMA, M. M. A.; VIDAL, M. S. Aspectos genéticos e moleculares de plantas submetidas ao déficit hídrico. **Revista Brasileira de Oleaginosas e Fibrosas**, v.10, n.1/2, p.1039-1072, 2006.

METZKER, M.L. Sequencing Technologies – the next generation. **Nature reviews**, v.11, p. 31-46, 2010.

MIRICA, L. M.; KLINMAN, J. P. The nature of O<sub>2</sub> activation by the ethylene-forming enzyme 1-aminocyclopropane-1-carboxylic acid oxidase. **Proceedings of the National Academy of Sciences**, v. 105, n. 6, p. 1814-1819, 2008.

MORAES, F. E. **Identificação em cana-de-açúcar de genes de resistência a *Mahanarva fimbriolata* (STÅL, 1854) (Hemiptera: Cercopidae)**. 2010. 67 f. Dissertação (Mestrado em Agronomia: Genética e Melhoramento de Plantas) - Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2010.

MORAES, V. R. A. **Imunodeteção em frações protéicas de cérebro de abelha *Apis mellifera* de proteínas associadas a vesículas sinápticas, miosina-v e CA2+/ Calmodulina quinase II (CaMKII)**. 2005, 119f. Dissertação ( Mestrado em Genética e Bioquímica) – Universidade Federal de Uberlândia, Uberlândia, 2005.

MOURA, D. S.; BERGEY, D. R.; RYAN, C. A. Characterization and localization of a wound-inducible type I serine-carboxypeptidase from leaves of tomato plants (*Lycopersicon esculentum* Mill.). **Planta**, v. 212, n. 2, p. 222-230, 2001.

MURATA, M. M. **Transcriptoma da interação de tangerina Satsuma (*Citrus unshiu*) e laranja doce Hamlin (*Citrus sinensis*) infectadas com *Xanthomonas citri* subsp. *citri*, agente causal do cancro cítrico**. 2013. 181 f. Dissertação (Mestrado em Agronomia: Genética e Melhoramento de Plantas) - Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2013.

NATARAJAN, S. S.; XU, C.; BAE, H.; CAPERNA, T. J.; GARRETT, W. M. Characterization of Storage Proteins in Wild (Glycine soja) and Cultivated (Glycine max) Soybean Seeds Using Proteomic Analysis. **Journal of agricultural and food chemistry**, v. 54, p. 3114 – 3120, 2006.

NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. 4<sup>a</sup> Ed, W. H. Freeman, New York, 2004.

NEUTELINGS, G. Lignin variability in plant cell walls: Contribution of new models. **Plant Science**, v. 181, p. 379-386, 2011.

NOBUTA, k.; VENU, R. C.; LU, C.; BELÓ, A.; VEMARAJU, K.; KULKARNI, K.; WANG, W.; PILLAY, M.; GREEN, P. J.; WANG, G. MEYERS, B. C. An expression atlas of rice mRNAs and small RNAs. **Nature Biotechnolgy**, v.25, n. 4, p. 473 - 477 , 2007.

NONGPIUR, R.; SONI, P.; KARAN, R.; SINGLA-PAREEK, S. L.; PAREEK, A. Histidine kinases in plants - Cross talk between hormone and stress responses. **Plant Signaling & Behavior**, v. 7, n. 10, p. 1230-1237, 2012.

OLIVEIRA, J. F. N. C. **Caracterização fisiológica e perfil de expressão gênica de cultivares de cana-de-açúcar (*Saccharum spp*) contrastantes para o déficit hídrico**. 2012, 327 f. Dissertação (Doutorado em Ciências) – Universidade de São Paulo, Escola Superior de Agricultura “ Luiz de Queiroz”, Piracicaba, 2012.

OSHLACK, A.; ROBINSON, M.D.; YOUNG, M.D. From RNA-Seq reads to differential expression results. **Genome Biology**, v. 11, p. 220 - 229, 2010.

PALHARES, A. C. **Mapeamento genético de marcadores AFLP e de retrotransposons em cana-de-açúcar (*Saccharum spp.*)**. 2010. 97f. Dissertação (Mestrado em Ciências – Área: Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2010.

PAPINI-TERZI, F.S.; ROCHA, F. R.; VENCIO, R. Z. N.; OLIVEIRA, K. C.; FELIX, J. D.; VICENTINI, R.; ROCHA, C. D.; SIMOES, A. C. Q.; ULIAN, E. C.; DI MAURO, S. M. Z.; DA SILVA, A. M.; PEREIRA, C. A. D.; MENOSSI, M.; SOUZA, G. M. Transcription profiling of signal transduction-related genes in sugarcane tissues. **DNA Research**, v. 12, p. 27-38, 2005.

PASSOS, G.A.S.; JORDAN, C.N.B. Projeto transcriptoma. **Biotecnologia Ciência e Desenvolvimento**, v.12, p.34-37, 2000.

PASZKIEWICZ, K.; STUDHOLME, D. J. De novo assembly of short sequence reads. **Briefings in Bioinformatics**, Oxford, v. 2, n. 5, p. 457-472, 2010.

PIMENTEL, C. **A relação da planta com a água**. Seropédica: EDUR, 2004. 191 p.

PINCELLI, R.P. **Tolerância à deficiência hídrica em cultivares de cana-de-açúcar avaliada por meio de variáveis morfofisiológicas**. 2010. 65f. Dissertação (Mestrado em Agronomia) – Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2010.

PINOTI, V. F. **Análise da expressão gênica de pequenos RNAs derivados de tRNAs (tRFs) em plantas**. 2012, 52f. Trabalho de conclusão de curso (Ciências Biológicas) - Universidade Estadual “Júlio de Mesquita Filho”, Botucatu, 2012.

PIZA, I. M. T.; LIMA, G. P. P.; BRASIL, O. G. Atividade de peroxidase e níveis de proteínas em plantas de abacaxizeiro micropropagadas em meio salino. **Revista Brasileira de Agrociência**, v. 9, n. 4, p. 361-366, 2003.

PRICE, A. H.; HENDRY, G. A. F. Iron-catalysed oxygen radical formation and its possible contribution to drought damage in nine native grasses and three cereals. **Plant Cell Environment**. v.14, p.477-484, 1991.

PROCANA, 2012 . **Sobre o setor sucroenergético** Disponível em: <<http://www.jornalcana.com.br/sobre-nos/>>. Acesso em 25 novembro 2014.

PURVIS, A. C.; SHEWFELT, R. L. Does the alternative pathway ameliorate chilling injury in sensitive plant tissues? **Physiologia Plantarum**, v.88, p.712-718. 1993.

QIAN, Y.; CHENG, Y.; CHENG, X.; JIANG, H.; ZHU, S.; CHENG, B. Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize. **Plant Cell Rep**, v. 30, p. 1347–1363, 2011.

QUEIROZ, R. J. B. **Resposta fisiológica e molecular de dois genótipos de milho à limitação hídrica**. 154f. Tese (Doutorado em Agronomia - Genética e Melhoramento de Plantas) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, 2010.

RAMAKRISHNAN, V.; MOORE, P. B. Atomic structures at last: the ribosome in 2000. **Current Opinion in Structural Biology**, v. 11, n. 2, p. 144–154, 2001.

RAMESH, P. Effect of different levels of drought during the formative phase on growth parameters and its relationship with dry matter accumulation in sugarcane. **Journal Agronomy & Crop Science**, v.185, p.83-89, 2000.

REA, P. A.; VATAMANIUK, O. K.; RIGDEN, D. J. Weeds, worms, and more. Papain's long-lost cousin, phytochelatin synthase. **Plant Physiology**, v. 136, n.1, p. 2463 – 2474, 2004.

REDDY, P. S; RAO, T. S. R. B.; SHARMA, K. K.; VADEZ, V. Genome-wide identification and characterization of the aquaporin gene family in *Sorghum bicolor* (L.). **Plant Gene**, v. 1, p. 18 – 28, 2015.

RIBEIRO, I. L. A. C. **Proteômica de cana-de-açúcar em condição de estresse hídrico**. 2010. 128 f. Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade Federal de Pernambuco, Recife, 2010.

ROBERTSON, G.; SCHEIN, J.; CHIU, R.; CORBETT, R.; FIELD, M.; JACKMAN, S. D.; MUNGALL, K.; LEE, S.; OKADA, H. M.; QIAN, J. Q.; GRIFFITH, M.; RAYMOND, A.; THIESSEN, N.; CEZARD, T.; BUTTERFIELD, Y. S.; NEWSOME, R.; CHAN, S. K.; SHE, R.; VARHOL, R.; KAMOH, B.; PHABHU, A. – L.; TAM, A.; ZHAO, Y. J.; MOORE, R. A.; HIRST, M.; MARRA, M. A.; JONES, S. J. M.; HOODLESS, P. A.; BIROL, I. *De novo* assembly and analysis of RNA-seq data. **Nature Methods**, v. 7, n. 11, p. 909-912, 2010.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, p. 139-140, 2010.

ROCHA, F. R.; PAPINI-TERZI, F. S; NISHIYAMA, M. Y.; VENCIO, R. Z.; VICENTINI, R.; DUARTE, R. D.; DE ROSA, V. E.; VINAGRE, F.; BARSALOBRES, C.; MEDEIROS, A. H.; RODRIGUES, F. A.; ULIAN, E. C.; ZINGARETTI, S. M.; GALBIATTI, J. A.; ALMEIDA, R. S.; FIGUEIRA, A. V.; HEMERLY, A. S.; SILVA-FILHO, M. C.; MENOSSI, M.; SOUZA, G. M. Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. **BMC Genomics**, v. 8, p. 71, 2007.

RODRIGUES, F.A.; GRAÇA, J.P.; LAIA, M.L.; NHANI-JR, A.; GALBIATI, J.A.; FERRO, M.I.T.; FERRO, J.A.; ZINGARETTI, S.M. Sugarcane genes differentially expressed during water deficit. **Biologia Plantarum**, v. 55, n.1, p. 43-53, 2011.

RODRIGUES, C. M. **Expressão diferencial de genes em laranja doce (*Citrus sinensis* L.) e em tangerina (*Citrus reticulata* blanco) em resposta à infecção por *Xylella fastidiosa*.** Tese (Doutorado em Genética) - Universidade Estadual Paulista, Instituto de Biociências de Botucatu, 2011.

RODRIGUES, F.A.; LAIA, M.L.; ZINGARETTI, S.M. Analysis of gene expression profiles under water stress in tolerant and sensitive sugarcane plants. **Plant Science**, v. 176, p. 286-302, 2009.

ROELOFS, D.; AARTS, M. G. M.; SCHAT, H.; VAN STRAALLEN, N. M. Functional ecological genomics to demonstrate general and specific responses to abiotic stress. **Functional Ecology**, v. 22, p. 8 – 18, 2008.

ROUDIER, F.; FERNANDEZ, A. G.; FUJITA, M.; HIMMELSPACH, R.; BORNER, G. H. H.; SCHINDELMAN, G.; SONG, S.; BASKIN, T. I.; DUPREE, P.; WASTENEYS, G. O.; BENFEY, P. N. COBRA, an Arabidopsis Extracellular Glycosyl-Phosphatidyl Inositol-Anchored Protein, Specifically Controls Highly Anisotropic Expansion through Its Involvement in Cellulose Microfibril Orientation. **The Plant Cell**, v. 17, p. 1749–1763, 2005.

ROUGEMONT, J.; AMZLLAG, A.; ISELI, C.; FARINELLI, L.; XENARIOS, I.; NAEF, F. Probabilistic base calling of Solexa sequencing data. **BMC Bioinformatics**, v. 9, p. 431-443, 2008. Disponível em: <<http://dx.doi.org/10.1186/1471-2105-9-431>>.

SACILOTO, R. F. Z. **Inserção do gene PR5K em cana-de-açúcar visando induzir resistência ao fungo da ferrugem *Puccinia melanocephala*.** 2003. 74f. Dissertação (Mestrado em Ciências – Área: Fisiologia e Bioquímica de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003.

SAMBROOK, J.; FRITCH, E. F.; MANIATS, T. Molecular cloning: a laboratory manual. **Cold Spring Harbor: Press**. 1989.

SANYAL, S. C.; LILJAS, A. The end of the beginning: structural studies of ribosomal proteins. **Current Opinion in Structural Biology**, v.10, n. 6, p. 633-636, 2000.

SARIKAS, A.; HARTMANN, T.; PAN, Z. The cullin protein family. **Genome Biology**, v. 12, p. 220, 2011.

SEESURIYACHAN, P.; KUNTIYA, A.; HANMOUNGJAI, P.; TECHAPUN, C.; CHAIYASO, T.; LEKSAWASDI, N. Optimization of Exopolysaccharide Overproduction by *Lactobacillus confusus* in solid State Fermentation under High Salinity Stress. **Biosci. Biotechol. Biochem.**, v. 76, n. 5, p. 912-917, 2012.

SEVIER, C. S.; KAISER, C. A. Formation and transfer of disulphide bonds in living cells. **Nature**, v. 3, p. 836 – 847, 2002.



SHAO, C.; DONG, A-W.; MA, X.; MENG, Y. Is Argonaute 1 the only effective slicer of small RNA-mediated regulation of gene expression in plants? **Journal of Experimental Botany**, v. 65, n. 22, p. 6293-6299, 2014.

SHENDURE, J.; JI, H. Next-generation DNA sequencing. **Nature Biotechnology**, v. 26, n. 10, p. 1135–1145, 2008.

SHI, Q-M.; YANG, X.; SONG, L.; XUE, H-W. Arabidopsis MSBP1 Is Activated by HY5 and HYH and Is Involved in Photomorphogenesis and Brassinosteroid Sensitivity Regulation. **Molecular Plant**, v. 4, n. 6, p. 1092–1104, 2011.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K.; SEKI, M. Regulatory network of gene expression in the drought and cold stress responses. **Current Opinion in Plant Biology**, v. 6, p. 410 – 417, 2003.

SHIRLEY, A. M.; CHAPPLE, C. Biochemical Characterization of Sinapoylglucose: Choline Sinapoyltransferase, a Serine Carboxypeptidase-like Protein That Functions as an Acyltransferase in Plant Secondary Metabolism. **The journal of Biological Chemistry**, v. 278, n. 22, p.19870–19877, 2003.

SILVA, M. A.; JIFON, J. L.; SILVA, J. A. G.; SHARMA, V. Use of physiological parameters as fast tools to screen for drought tolerance in sugarcane. **Brazilian Journal of Plant Physiology**, v.19, n.3, p.193-201, 2007.

SILVA, M. D.; SILVA, R. L. O.; NETO, J. R. C. F.; GUIMARÃES, A. C. R.; VEIGA, D. T.; CHABREGAS, S. M.; BURNQUIST, W. L.; KAHL, G.; BENKO-ISEPPON, A. M.; KIDO, E. A. Expression Analysis of Sugarcane Aquaporin Genes under Water Deficit. **Journal of Nucleic Acids**, v. 2013, p. 1-14, 2013.

SILVA, P. B. **Aspectos fisiológicos de seis genótipos de cana-de-açúcar submetidos a estresse hídrico**. 2010. 96p. Dissertação (Mestrado em Agronomia) – Centro de Ciências Agrárias, Universidade Federal de Alagoas, 2010.

SINGH, S.; RAO, P.N.G. Varietal differences in growth characteristics in sugarcane. **The Journal of Agricultural Science**, v.108, p.245-247, 1987.

SMIRNOFF, N. The role of active oxygen in the response of plants to water deficit and desiccation. **New Phytologist**. v.125, p.27-58, 1993.

SOUZA, C. B. **Biotechnologia de cana-de-açúcar (*Saccharum spp.*) para tolerância a estresse hídrico**. 2012. 156 f. Dissertação (Doutorado em Genética e Biologia Molecular) – Universidade Estadual de Campinas, Campinas, 2012.

STOCK, A. M.; ROBINSON, V. L.; GOUDREAU, P. N. Two-component signal transduction. **Annual Review Biochemistry**, v. 69, p. 183–215, 2000.

SUGIHARTO, B.; ERMAWATI, N.; MORI, H.; AOKI, K.; YONEKURA-SAKAKIBARA, K.; YAMAYA, T.; SUGIYAMA, T.; SAKAKIBARA, H. Identification and characterization of a gene encoding drought-inducible protein localizing in the bundle sheath cell of sugarcane. **Plant & Cell Physiology**, v. 43, p. 350-354, 2002.

SUGIHARTO, B. Biochemical and Molecular Studies on Sucrose-phosphate Synthase and Drought Inducible-protein in Sugarcane (*Saccharum officinarum*). **Journal ILMU DASAR**, v. 5, p. 62-67, 2004.

SURGET-GROBA, Y.; MONTOYA-BURGOS, J. I. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. **Genome Research**, v. 20, p. 1432-1440, 2010.

TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. Porto Alegre: Artmed, 2009, 843 p.

TELLES, B. R. **Desempenho morfofisiológico de cultivares de cana-de-açúcar submetidas ao estresse hídrico prolongado**. 2013. 59 f. Monografia (Ciências Biológicas) - Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2013.

TESTER, M.; BACIC, A. Abiotic stress tolerance in grasses. From model plants to crop plants. **Plant Physiology**, v. 137, p. 791-793, 2005.

UMATE, P.; TUTEJA, R.; TUTEJA, N. Genome-wide analysis of helicase gene family from rice and Arabidopsis: a comparison with yeast and human. **Plant Molecular Biology**, v. 73, p. 449-465, 2010.

ÚNICA- União da Indústria de Cana-de-Açúcar. **Setor sucroenergético – Mapa de Produção**. 2014. Disponível em: < <http://www.unica.com.br/mapa-da-producao/>> Acesso em: 15 dez. 2014.

UNIPROTKB- UniProt Knowledgebase. **UniProtKB/Swiss-Prot protein knowledgebase release 2015\_05 statistics**. Disponível em: < <http://web.expasy.org/docs/relnotes/relstat.html>> Acesso em 22 abril 2015.

VAN VELLEET, A.H. Next generation sequencing of microbial transcriptomes: challenges and opportunities. **FEMS Microbiology Letters**, v. 302, p. 1-7, 2010. Disponível em: <<http://dx.doi.org/10.1111/j.1574-6968.2009.01767.x>>.

VANHOLME, R.; DEMEDTS, B.; MORREEL, K.; RALPH, J.; BOERJAN, W. Lignin biosynthesis and structure. **Plant Physiology**, v. 153, p. 895-905, 2010.

VANTINI, J. S. **Expressão gênica diferencial por cDNA-AFLP em raízes de cana-de-açúcar submetida ao déficit hídrico**. 2013. 105 f. Dissertação (Doutorado em Agronomia: Genética e Melhoramento de Plantas) - Universidade Estadual Paulista - UNESP, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, 2013.

VANTINI, J. S.; DEDEMO, G. C.; JOVINO GIMENEZ, D. F.; FONSECA, L. F. S.; TEZZA, R. I. D.; MUTTON, M. A.; FERRO, J. A.; FERRO, M. I. T. Differential gene expression in drought-tolerant sugarcane roots. **Genetics and Molecular Research**, v. 14, n. 2, p. 7196-7207, 2015.

VETTORE A. L.; DA SILVA, F. R.; KEMPER, E.; ARRUDA, P. The libraries that made SUCEST. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1-4, p. 1-7, 2001.

VETTORE, A. L.; DA SILVA, F. R.; KEMPER, E. L.; SOUZA, G. M.; DA SILVA, A. M.; FERRO, M. I., HENRIQUE-SILVA, F.; GIGLIOTI, E. A.; LEMOS, M. V.; COUTINHO, L. L.; NOBREGA M. P.; CARRER, H.; FRANÇA, S. C.; BACCI JÚNIOR, M.; GOLDMAN, M. H.; GOMES, S. L.; NUNES, L. R.; CAMARGO, L. E.; SIQUEIRA, W. J.; VAN SLUYS, M. A.; THIEMANN, O. H.; KURAMAE, E. E.; SANTELLI, R. V.; MARINO, C. L.; TARGON, M. L.; FERRO, J. A.; SILVEIRA, H. C.; MARINI, D. C.; LEMOS, E. G.; MONTEIRO-VITORELLO, C. B.; TAMBOR, J. H.; CARRARO, D. M.; ROBERTO, P. G.; MARTINS, V. G.; GOLDMAN, G. H.; DE OLIVEIRA, R. C.; TRUFFI, D.; COLOMBO, C. A.; ROSSI, M.; DE ARAUJO, P. G.; SCULACCIO, S. A.; ANGELLA, A.; LIMA, M. M.; DE ROSA JÚNIOR, V. E.; SIVIERO, F.; COSCRATO, V. E.; MACHADO, M. A.; GRIVET, L.; DI MAURO, S. M.; NOBREGA, F. G.; MENCK, C. F.; BRAGA, M. D.; TELLES, G. P.; CARA, F. A.; PEDROSA, G.; MEIDANIS, J.; ARRUDA, P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, v.13, p. 2725–2735, 2003.

VINCENT, D.; LAPIERRE, C.; POLLET, B.; CORNIC, G.; NEGRONI, L.; ZIVY, M. Water deficits affect caffeate O-methyltransferase, lignification, and related enzymes in maize leaves. A proteomic investigation. **Plant Physiology**, v.137, p. 949-960, 2005.

WANG, R.; GAO, F.; GUO, B-Q.; HUANG, J-C.; WANG, L.; ZHOU, Y-J. Short-Term Chromium-Stress-Induced Alterations in the Maize Leaf Proteome. **International Journal of Molecular Sciences**, v. 14, p. 11125 – 11144, 2013.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Review Genetics**, v.10, p. 57–63, 2009. Disponível em: <<http://dx.doi.org/10.1038/nrg2484>>.

WATSON, J. D.; BAKER, T. A.; BELL, S. P.; GANN, A.; LEVINE, M.; LOSICK, R. **Biologia Molecular do Gene**. 5.<sup>a</sup> ed. Porto Alegre, Artmed, 2006. 728p.

WILHELM, B. T.; MARGUERAT, S.; GOODHEAD, I.; BÄHLER, J. Defining transcribed regions using RNA-Seq. **Nature Protocols**, v. 5, p. 255-266, 2010.

WILSON, I. B. H. Glycosylation of proteins in plants and invertebrates. **Structural Biology**, v. 12, p. 569–577, 2002.

XIE, F.; STEWART JR, C. N.; TAKI, F. A.; HE, Q.; LIU, H.; ZHANG, B. High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. **Plant Biotechnology Journal**, v. 12, p. 354–366, 2014.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Organization of cis-acting regulatory elements in osmotic and cold stress responsive promoters. **Trends in Plant Science**, v. 10, p. 88-94, 2005.

YANG, L.; WANG, C. C.; GUO, W. D.; LI, X. B.; LU, M.; YU, C. L. Differential expression of cell wall related genes in the elongation zone of rice roots under water deficit. **Russian Journal of Plant Physiology**, v. 53, p. 390-395, 2006.

YANG, X-H.; XU, Z-H.; XUE, H-W. Arabidopsis Membrane Steroid Binding Protein 1 Is Involved in Inhibition of Cell Elongation. **The Plant Cell**, v. 17, p. 116–131, 2005.

YE, J.; FANG, L.; ZHENG, H.; ZHANG, Y.; CHEN, J.; ZHANG, Z.; WANG, J.; LI, S.; LI, R.; BOLUND, L.; WANG, J. WEGO: a web tool for plotting GO annotations. **Nucleic Acids Research**, v. 34, p. W293–W297, 2006.

YIGIT, E.; BATISTA, P. J.; BEI Y.; PANG, K. M.; CHEN, C. C.; TOLIA, N. H.; JOSHUA-TOR, L.; MITANI, S.; SIMARD, M. J.; MELLO, C. C. Analysis of the *C. elegans* Argonaute family reveals that distinct argonautes act sequentially during RNAi. **Cell**, v.127, n. 4, p. 747–757, 2006.

YOSHIMURA, K.; MASUDA, A.; KUWANO, M.; YOKOTA, A.; AKASHI, K. Programmed proteome response for drought avoidance/tolerance in the root of a C-3 xerophyte (wild watermelon) under water deficits. **Plant Cell Physiology**, v. 49, p. 226-241, 2008.

ZDOBNOV, E. M.; APWEILER, R. InterProScan--an integration platform for the signature-recognition methods in InterPro. **Bioinformatics**, v. 17, p. 847-848, 2001.

ZERBINO, D. R.; BIRNEY, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. **Genome Research**, v. 18, p. 821–829, 2008.

ZHANG, S.; QI, Y.; LIU, M.; YANG, C. SUMO E3 Ligase AtMMS21 Regulates Drought Tolerance in *Arabidopsis thaliana*. **Journal of Integrative Plant Biology**, v. 55, n. 1, p. 83-95, 2013.