

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**IDENTIFICAÇÃO DE SNPs CANDIDATOS RELACIONADOS
À TOLERÂNCIA AO ESTRESSE HÍDRICO EM
CANA-DE-AÇÚCAR**

Francini Ludmila Tulini
Engenheira Agrônoma

2020

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**IDENTIFICAÇÃO DE SNPs CANDIDATOS RELACIONADOS
À TOLERÂNCIA AO ESTRESSE HÍDRICO EM
CANA-DE-AÇÚCAR**

Discente: Francini Ludmila Tulini

Orientadora: Profa. Dra. Maria Inês Tiraboschi Ferro

Coorientador: Prof. Dr. Daniel Guariz Pinheiro

Coorientadora: Dra. Daniele Fernanda Jovino Gimenez

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).

2020

T917i

Tulini, Francini Ludmila

Identificação de SNPs candidatos relacionados à tolerância ao estresse hídrico em Cana-de-açúcar / Francini Ludmila Tulini. -- Jaboticabal, 2020

90 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientadora: Maria Inês Tiraboschi Ferro

Coorientador: Daniel Guariz Pinheiro

1. Transcriptoma. 2. Polimorfismo de nucleotídeo único. 3. Cana-de-açúcar. 4. Desidratação (Hídrica). I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: IDENTIFICAÇÃO DE SNPs CANDIDATOS RELACIONADOS À TOLERÂNCIA AO ESTRESSE HÍDRICO EM CANA-DE-AÇÚCAR

AUTORA: FRANCINI LUDMILA TULINI

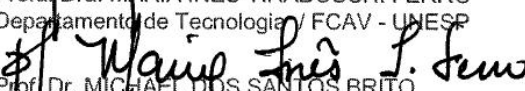
ORIENTADORA: MARIA INES TIRABOSCHI FERRO

COORIENTADORA: DANIELE FERNANDA JOVINO GIMENEZ

COORIENTADOR: DANIEL GUARIZ PINHEIRO

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


Prof.ª Dra. MARIA INES TIRABOSCHI FERRO
Departamento de Tecnologia / FCAV - UNESP


Prof. Dr. MICHAEL DOS SANTOS BRITO (VIDEOCONFERÊNCIA)
Instituto de Ciência e Tecnologia-Unifesp / São José dos Campos/SP


Dr. ANTONIO NHANI JÚNIOR (VIDEOCONFERÊNCIA)
EMBRAPA Informática Agropecuária / Campinas/SP

Jaboticabal, 19 de junho de 2020

DADOS CURRICULARES DA AUTORA

Francini Ludmila Tulini - nascida em 24 de abril de 1992, no município de Ribeirão Preto, São Paulo, filha de Edina Pereira da Motta Tulini e Luiz Roberto Tulini. Formou-se em Engenharia Agrônômica pela Faculdade de Ciências Agrárias e Veterinárias - Unesp/Jaboticabal em Janeiro de 2017. Em março de 2018 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, sob a orientação da Profa. Dra. Maria Inês Tiraboschi Ferro, o que resultou no presente trabalho.

*“ Nada na vida deve ser temido, somente compreendido.
Agora é hora de compreender mais para temer menos. ”*

Marie Curie

AGRADECIMENTOS

À Deus, por ter permitido que tudo isso acontecesse ao longo de toda minha vida, e não somente nestes anos como universitária, mas que em todos os momentos é o maior mestre que alguém pode conhecer.

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 Profa. Dra. Maria Inês Tiraboschi Ferro, pela orientação, ensinamentos, confiança, carinho, incentivo e amizade durante todo desenvolvimento desse trabalho. Muito obrigada!

Ao Prof. Dr. Daniel Guariz Pinheiro, pela coorientação, correções e ajuda nas análises de Bioinformática.

À Dra. Daniele Fernanda Jovino Gimenez, pela coorientação, ensinamentos, auxílios e preciosas dicas nos trabalhos de bancada.

Aos componentes da banca de qualificação: Profa. Dra. Maria Inês Tiraboschi Ferro, Profa. Dra. Luciana Rossini Pinto e Dra. Poliana Fernanda Giachetto agradeço pela participação e contribuição.

Aos componentes da comissão examinadora: Antonio Nhani Júnior e Michael dos Santos Brito, agradeço pela disponibilidade e participação.

Ao Lucas Amoroso Lopes de Carvalho, que foi de fundamental importância para concluir os dados e análises de bioinformática, meu mais sincero obrigada!

À Profa. Dra. Lúcia Maria Carareto Alves pela oportunidade de realizar o estágio docência, pelo carinho e pelos ensinamentos.

Ao Prof. Dr. Jesus Aparecido Ferro pelas contribuições e disponibilidade no laboratório.

À minha querida amiga Helô, por ser essa pessoa incrível que sempre esteve por perto para ajudar no que fosse preciso. Muito obrigada, amiga!

Aos queridos amigos do Laboratório de Bioquímica e Biologia Molecular e do Laboratório de Bioinformática: Amanda, Helô, Dani Konrad, Dani Gimenez, Flávia,

Helen, Jú Vantini, Iashilei, Luana, Lucas, Luis, Michelli e Rafael Marini, agradeço pela convivência, amizade e pelos momentos de descontração.

Às meninas do Centro de Recursos Biológicos e Genômicos (CREBIO): Dani, Agda, Mariza e Célia, agradeço pela colaboração e disponibilidade em ajudar sempre que possível.

A todos os alunos, funcionários e professores do Departamento de Tecnologia, que de alguma maneira contribuíram para a realização desse trabalho.

Aos meus amigos do coração, Maíra, Vanessa e Fabio, pelo carinho, força e incentivo nos momentos difíceis.

Aos meus queridos pais Edina e Beto, pelo amor, incentivo e apoio incondicional.

Aos meus irmãos Priscila e Fabrício, por estarem sempre ao meu lado.

Ao meu namorado e melhor amigo Diego, por acreditar no meu sucesso, mesmo quando tudo era incerto.

À Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento da pesquisa;

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Gratidão a todos!

SUMÁRIO

| | |
|--|------------|
| RESUMO | iii |
| ABSTRACT | iv |
| LISTA DE ABREVIATURAS | v |
| LISTA DE TABELAS | vi |
| LISTA DE FIGURAS | vii |
| 1. INTRODUÇÃO | 1 |
| 2. REVISÃO DE LITERATURA | 5 |
| 2.1. Cana-de-açúcar: origem, aspectos gerais e importância econômica | 5 |
| 2.2. Estresse hídrico em plantas..... | 8 |
| 2.3. Mecanismos de resistência ao estresse | 11 |
| 2.4. Estudo do transcriptoma via RNA-Seq | 13 |
| 2.5. Marcadores Moleculares do tipo SNP (<i>Single Nucleotide Polymorphisms</i>) | 18 |
| 2.6. Estudos fisiológicos, genéticos e moleculares em cana-de-açúcar submetida a prolongada limitação hídrica | 22 |
| 3. MATERIAL E MÉTODOS | 24 |
| 3.1. Experimento em casa de vegetação, delineamento experimental e material vegetal | 24 |
| 3.2. Extração, quantificação e análise da integridade do RNA total | 26 |
| 3.3. Construção das bibliotecas de cDNA e sequenciamento via RNA-Seq na plataforma Illumina | 26 |
| 3.4. Processamento dos dados para identificação de SNPs candidatos | 26 |
| 3.4.1. Pré-processamento das leituras e montagem <i>de novo</i> dos transcritos | 27 |
| 3.4.2. Análise de expressão diferencial por cultivar, nível e tempo de estresse | 28 |
| 3.4.3. Anotação dos transcritos | 28 |
| 3.4.4. Alinhamento e detecção de bases variantes | 29 |
| 3.4.5. Predição dos efeitos das variações | 30 |
| 3.4.6. Seleção de variantes marcadores | 30 |
| 3.4.7. Categorização funcional e análise de enriquecimento com base nos termos de ontologia gênica..... | 31 |

| | |
|---|-----------|
| 4. RESULTADOS E DISCUSSÃO | 32 |
| 4.1. Qualidade do material, sequenciamento das bibliotecas de cDNA via RNA-Seq e processamento das leituras brutas..... | 32 |
| 4.2. Montagem <i>de novo</i> | 32 |
| 4.3. Anotação dos transcritos | 32 |
| 4.4. Identificação dos SNPs putativos | 33 |
| 4.5. Categorização funcional e análise de enriquecimento com base nos termos de ontologia gênica..... | 39 |
| 4.6. Genes associados a SNPs putativos relacionados à resposta ao estresse hídrico | 41 |
| 4.6.1. Proteínas quinases dependentes de cálcio (CDPKs), Serina/Treonina quinases (Ser/Tre) e Receptores do tipo quinase com repetições ricas em leucina (LRR-RPK)..... | 45 |
| 4.6.2. Fatores de Transcrição ABF, ARF, MYB e Zinc Finger | 46 |
| 4.6.2.1. Fator de ligação do elemento responsivo ao ABA (ABF) | 47 |
| 4.6.2.2. Fator de resposta a auxina (ARF)..... | 48 |
| 4.6.2.3. Fatores de transcrição MYB | 49 |
| 4.6.2.4. Fatores de Transcrição Dedos de Zinco (“Zinc Finger”) | 50 |
| 4.6.2.5. Fatores de transcrição de estresse térmico (HSF) e Proteínas de Choque Térmico (HSP) | 51 |
| 4.6.3. Glutathiona S-transferase | 52 |
| 4.6.4. Universal Stress Protein | 54 |
| 5. CONCLUSÕES | 56 |
| 6. REFERÊNCIAS | 58 |

IDENTIFICAÇÃO DE SNPs CANDIDATOS RELACIONADOS À TOLERÂNCIA AO ESTRESSE HÍDRICO EM CANA-DE-AÇÚCAR

RESUMO - A cana-de-açúcar é uma das culturas mais importantes do Brasil sob o ponto de vista socioeconômico. O aumento da demanda pelos subprodutos desta cultura, principalmente o açúcar e o etanol, tem atraído a canavieira para regiões do país que apresentam longos períodos de deficiência hídrica, refletindo em perda de produtividade. Assim, uma das principais estratégias para contornar esse problema é o desenvolvimento de cultivares tolerantes à seca. Entretanto, o alto nível de complexidade genética da cana-de-açúcar é um desafio para a aplicação de ferramentas moleculares no melhoramento dessa cultura. As novas tecnologias de sequenciamento de alto desempenho são valiosas ferramentas para o estudo de sequências genômicas e de transcriptomas, auxiliando na identificação de regiões do genoma que estejam relacionadas com características agrônômicas de interesse para o melhoramento. Com isso, o presente estudo teve como objetivo identificar SNPs (polimorfismos de nucleotídeo único) candidatos dentro de genes relacionados com respostas ao estresse hídrico para que possam ser utilizados como marcadores moleculares na obtenção de cultivares de cana-de-açúcar tolerantes ao estresse hídrico. Para isso, utilizamos duas cultivares de cana-de-açúcar contrastantes quanto ao déficit hídrico: a SP81-3250 (tolerante) e a RB85-5453 (sensível). Essas plantas foram submetidas a três potenciais hídricos do solo: sem restrição hídrica, déficit hídrico moderado e severo, aplicados 60 dias após o plantio. Em seguida foram avaliadas em três épocas distintas: aos 30, 60 e 90 dias após aplicação dos tratamentos, sendo um dos primeiros estudos a avaliar os efeitos do déficit hídrico prolongado em cana-de-açúcar. A partir do sequenciamento via RNA-Seq obtivemos um transcriptoma conjunto para as duas cultivares. Os sítios variantes (sobretudo os SNPs) foram detectados com o “software” FreeBayes e os efeitos funcionais preditos foram anotados com o programa SnpEff. Após filtragem dos dados a fim de remover falsos positivos, foram obtidas 5.866 variações exclusivas da cultivar tolerante, 9.797 variações exclusivas da cultivar sensível e 10.349 variações em comum para ambas as cultivares. Alguns dos SNPs exclusivos da cultivar tolerante foram encontrados em genes relacionados com a resposta da planta ao estresse hídrico, como fatores de transcrição ABF, ARF, MYB e “Zinc Finger”, proteínas quinases, proteínas de choque térmico, proteína universal do estresse e enzimas de detoxificação como a Glutathione S-transferase. Por essa razão, acreditamos que polimorfismos específicos dessas regiões de resposta estejam associados com a capacidade das cultivares contrastantes lidarem com o estresse hídrico.

Palavras-chave: RNA-Seq, Transcriptoma, Poliploides, Marcadores moleculares, Tolerância a seca, *Saccharum* spp.

IDENTIFICATION OF SNPs CANDIDATES RELATED TO WATER STRESS TOLERANCE IN SUGARCANE

ABSTRACT - Sugarcane is one of the most important crops in Brazil with regard to the economic aspects. The increasing demand for sugarcane by-products, especially sugar and ethanol, has promoted the sugarcane cultivation in regions with long periods of water deficiency, resulting in loss of productivity. One of the main strategies to overcome this problem is the development of drought-tolerant cultivars. However, sugarcane has a high level of genetic complexity, which is a challenge for applying molecular tools to improve this crop. The new high-performance sequencing technologies are valuable tools for the study of genomic and transcriptome sequences, which allows the identification of regions of the genome related to the improvement of agronomic characteristics. Thus, the present study aimed to identify SNPs (single nucleotide polymorphisms) candidates within genes related to water stress responses, which may be used as molecular markers to obtain sugarcane cultivars that are tolerant to water stress. For this, it was used two contrasting sugarcane cultivars in terms of water deficit: SP81-3250 (tolerant) and RB85-5453 (sensitive). These cultivars were submitted to three soil water potentials: without water restriction, or with moderate and severe water deficit, which were applied 60 days after planting. Next, they were evaluated at 30, 60, and 90 days after the application of treatments. This study was one of the first to assess the effects of prolonged water deficit in sugarcane. RNA-Seq sequencing was used to obtain a joint transcriptome for the two cultivars. Variant sites (especially SNPs) were detected with the FreeBayes software, and the predicted functional effects were noted with the SnpEff software. After filtering the data to remove false positives, it was detected 5,866 exclusive variations of the tolerant cultivar, 9,797 exclusive variations of the sensitive cultivar, and 10,349 variations in common for both cultivars. Some of the SNPs specific to the tolerant cultivar were found in genes related to the plant response to water stress, such as ABF transcription factors, ARF, MYB and Zinc Finger, protein kinases, heat shock proteins, universal stress protein and detoxification enzymes such as glutathione S-transferase. For this reason, we believe that specific polymorphisms in these response regions are associated with the ability of contrasting cultivars to deal with water stress.

Keywords: RNA-Seq, Transcriptome, Polyploids, Molecular markers, Drought tolerance, *Saccharum* spp.

LISTA DE ABREVIATURAS

| | |
|--------|--|
| ABA | Ácido abscísico |
| ABF | Fator de ligação do elemento responsivo ao ABA |
| AIA | Ácido indolacético |
| ARF | Fator de resposta a auxina |
| CDPK | Proteína quinase dependente de cálcio |
| EROs | Espécies reativas de oxigênio |
| FT | Fator de transcrição |
| GST | Glutathiona S-Transferase |
| HSF | Fator de transcrição de estresse térmico |
| HSP | Proteína de choque térmico |
| Indels | Inserções e deleções |
| SAM | Seleção assistida por marcadores |
| SNP | Polimorfismo de nucleotídeo único |
| USP | Proteína universal de estresse |
| ZF | “Zinc Finger” ou Dedos de zinco |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1. Características agronômicas das cultivares utilizadas neste estudo. Adaptado de Marin (2019)..... | 25 |
| Tabela 2. Anotação dos transcritos por referência..... | 33 |
| Tabela 3. Classificação das variações em dados brutos e filtrados nas cultivares sensível e tolerante à estresse hídrico de acordo com a magnitude predita de seu efeito. Os sítios variantes foram classificados pelo SnpEff nas categorias de impacto alto, moderado, baixo e modificador..... | 37 |
| Tabela 4. SNPs candidatos presentes em genes de resposta ao estresse hídrico. ID: identificação do gene; POS: posição do SNP na sequência; REF: nucleotídeo de referência; ALT: nucleotídeo alterado pelo SNP..... | 42 |

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. Estágios de desenvolvimento da cana-de-açúcar. Adaptado de Gascho e Shih (1983)..... | 6 |
| Figura 2. Mapa de produção da cana-de-açúcar no Brasil (Unica, 2019)..... | 7 |
| Figura 3. Representação esquemática dos mecanismos de defesa desencadeados pelas plantas em resposta à estresses abióticos. Adaptado de Wang et al. (2003)..... | 12 |
| Figura 4. Esquema geral de um experimento utilizando a técnica de RNA-Seq na plataforma Illumina/Solexa. A etapa 1 engloba a preparação da amostra a partir do RNA extraído do organismo de estudo. A etapa 2 consiste na construção de bibliotecas de cDNA a partir das amostras de RNAm purificado. Na etapa 3 ocorre a clonagem “in vitro” para amplificação das amostras através da técnica de PCR. A etapa 4 é o sequenciamento das bibliotecas de cada amostra na plataforma Illumina/Solexa. Por último, a etapa 5 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. Adaptado de Jain (2012)..... | 16 |
| Figura 5. Condução do experimento em casa de vegetação após aplicação das três distintas tensões hídricas (sem restrição hídrica, restrição hídrica moderada e restrição hídrica severa) – A) Cinco dias após aplicação dos tratamentos, B) Sessenta dias após aplicação dos tratamentos e C) Noventa dias após a aplicação dos tratamentos. Imagem retirada de Telles (2016)..... | 25 |
| Figura 6. Resumo da metodologia utilizada para as análises de bioinformática após sequenciamento via RNA-Seq..... | 27 |
| Figura 7. Número total de variações encontradas exclusivamente em cada cultivar (Sensível e Tolerante) e em comum para ambas, considerando apenas aquelas que ocorrem em mais de metade das réplicas biológicas de cada condição e destas, apenas aquelas que ocorrem em mais de metade das bibliotecas da cultivar relacionada..... | 35 |
| Figura 8. Distribuição das diferentes categorias de polimorfismos nas duas cultivares de cana-de-açúcar: (A) aqueles que ocorrem exclusivamente na cultivar sensível; (B) aqueles que ocorrem exclusivamente na cultivar tolerante. SNP - polimorfismo de nucleotídeo único; MNP - polimorfismo de múltiplos nucleotídeos; INS - inserção; DEL - deleção e COMPLEX - variações complexas..... | 35 |
| Figura 9. Distribuição dos polimorfismos de acordo com a posição que ocorrem no gene. (A) cultivar sensível; (B) cultivar tolerante..... | 36 |
| Figura 10. Efeitos preditos dos polimorfismos nos genes, classificados em impacto alto segundo classificação de Cingolani et al. (2012)..... | 38 |

- Figura 11.** Efeitos preditos dos polimorfismos nos genes, classificados em impacto moderado segundo classificação de Cingolani et al. (2012)..... 39
- Figura 12.** Categorização funcional e análise de enriquecimento com base nos termos de ontologia gênica realizada através do programa WEGO. As barras verdes representam genes com SNPs exclusivos da cultivar sensível e barras azuis são aqueles genes com SNPs exclusivos da cultivar tolerante..... 40

1. INTRODUÇÃO

O Brasil é o maior produtor mundial de cana-de-açúcar, responsável pela produção de cerca de 620 milhões de toneladas na safra 2018/2019. As características favoráveis do país, como condições edafoclimáticas propícias à cultura e extensas áreas de cultivo, fazem do Brasil um importante exportador dessa “commodity” e, conseqüentemente, torna grande a participação da cana-de-açúcar no agronegócio brasileiro, representando 2% do Produto Interno Bruto (PIB) brasileiro com a produção de açúcar e etanol e gerando mais de 800.000 empregos no interior do País. (Conab, 2019).

A produção de cana-de-açúcar no Brasil foi impulsionada com o surgimento do Programa Nacional do Álcool (Proálcool), criado pelo governo brasileiro em 1975 após a crise do petróleo elevar os custos com importação. O programa tinha como objetivo incentivar a substituição de combustíveis derivados do petróleo por etanol e, em menos de cinco anos aumentou a produção de pouco mais de 300 milhões de litros para 11 bilhões de litros. A partir de então a cana-de-açúcar ganhou destaque no país, atraindo investimentos em programas de melhoramento e novas tecnologias para a indústria e para o campo (Matsuoka et al., 2009).

Com aumento da demanda mundial por biocombustíveis as áreas canavieiras estão avançando para regiões onde períodos de secas severas são comuns, como o oeste e noroeste do estado de São Paulo, o triângulo mineiro, o leste do Mato Grosso do Sul, Goiás, Tocantins, Maranhão e oeste da Bahia. Estas regiões apresentam períodos de déficit hídrico bastante acentuados quando comparadas às regiões tradicionalmente ocupadas pela cultura da cana-de-açúcar. O déficit hídrico é um dos principais fatores que limitam a produção da cana-de-açúcar. Este é um problema a ser superado, uma vez que o ciclo da cana-de-açúcar é considerado longo (12 a 18 meses do plantio até a colheita) e o uso de irrigação é uma alternativa pouco utilizada e muito onerosa. Com isto os produtores de cana não têm como fugir das alterações climáticas sobre a cultura no decorrer de no mínimo um ano (Castro et al., 2010).

No Brasil, os impactos previstos das mudanças climáticas sobre a precipitação atmosférica e os déficits de umidade do solo nas fases críticas do crescimento das culturas reforçam a necessidade urgente de realizar análises mais detalhadas das

áreas prioritárias de cultivo, com o objetivo de desenvolver cultivares mais tolerantes à seca, combinadas a solos férteis e a estratégias de manejo da água para diminuir os efeitos previstos (Assad et al., 2013).

O ciclo da cana-de-açúcar apresenta quatro estádios fenológicos distintos: germinação, perfilhamento, crescimento intenso e maturação (Gascho e Shih, 1983). As fases de perfilhamento e crescimento intenso, 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 nestas fases aproximadamente 70 a 80% da produção de cana-de-açúcar é constituída (Singh e Rao, 1987; Smit e Singels, 2006; Zhao et al., 2010).

A diminuição do teor de água no solo afeta alguns processos morfofisiológicos como transporte fotoquímico de elétrons, fotossíntese e partição de fotoassimilados, resultando em redução das trocas gasosas, redução do tamanho das plantas e da área foliar, e conseqüentemente, redução da produtividade (Taiz et al., 2017; Endres et al., 2010), cuja irreversibilidade vai depender do genótipo, da duração, da severidade e do estágio de desenvolvimento da planta (Pimentel, 2004).

É preciso destacar que a deficiência hídrica é um estresse multidimensional, que causa vários efeitos fisiológicos e bioquímicos nas plantas (Jamaux et al., 1997) e a tolerância a essa condição é uma característica multigênica (Rodrigues et al., 2009; Rodrigues et al., 2011), de extrema dificuldade em identificar características únicas que possam ser utilizadas para a seleção de cultivares tolerantes.

O melhoramento genético clássico da cana-de-açúcar baseia-se na seleção e clonagem de genótipos superiores de populações segregantes obtidas por meio de cruzamentos biparentais ou multiparentais (Matsuoka et al., 2005). Embora os programas de melhoramento de cana-de-açúcar já possuam metodologias bem estabelecidas para seleção e liberação de novos cultivares, o tempo de obtenção de um novo cultivar, que ocorre em média de 10 a 15 anos, ainda é considerado longo (Raboin et al., 2008).

Dessa forma, uma alternativa para superar o problema da deficiência hídrica é lançar mão de técnicas de melhoramento molecular e biotecnologia para auxiliar no desenvolvimento de cultivares tolerantes à seca que se adaptem às novas regiões de cultivo da cana-de-açúcar, garantindo produtividade igual ou superior às regiões

tradicionais de cultivo. Marcadores moleculares, como os polimorfismos de nucleotídeo único (SNPs - *Single Nucleotide Polymorphism*) por exemplo, podem ser usados para acelerar o processo de obtenção de novas cultivares, através da seleção de indivíduos que serão utilizados em cruzamentos dirigidos (Cardoso-Silva et al., 2014).

Os SNPs são polimorfismos específicos que ocorrem em uma única posição no genoma, em um único nucleotídeo. Esse tipo de marcador tem uma ampla distribuição no genoma da maioria das espécies de plantas, inclusive em cana-de-açúcar. Na literatura é possível encontrar diversos estudos genéticos que utilizam esse tipo de marcador em cana-de-açúcar (Cordeiro et al., 2006; Garcia et al., 2013; Cardoso-Silva et al., 2014; Costa et al., 2016).

A maioria dos estudos envolvendo déficit hídrico severo em cana-de-açúcar para identificação de características relacionadas à tolerância ou suscetibilidade ao estresse são realizados com curto período de aplicação do estresse (Rodrigues et al., 2009; Gupta et al., 2010; Rodrigues et al., 2011). Entretanto, essa estratégia pode levar a resultados que podem não corresponder às situações reais de campo, onde períodos de estiagem podem durar vários meses.

Com isso, nosso grupo de pesquisa iniciou um estudo a fim de avaliar a resposta de plantas de cana-de-açúcar ao déficit hídrico prolongado. O sequenciamento de RNA (RNA-Seq) de folhas de cana-de-açúcar permitiu avaliar o comportamento fisiológico (Telles et al., 2019) e o perfil de transcrição (Belesini et al., 2017; Konrad, 2019) de duas cultivares contrastantes, uma tolerante (SP81-3250) e outra sensível (RB85-5453) ao déficit hídrico, em resposta ao tempo e severidade do estresse aplicado.

A estratégia de sequenciamento em larga-escala do transcriptoma (RNA-seq) além de possibilitar a obtenção de informações a respeito dos genes expressos, também possibilita a identificação de SNPs putativos nas regiões transcritas, os quais podem ser utilizados na detecção de expressão alelo-específica e também como marcadores moleculares. Através dessa tecnologia é possível identificar SNPs exclusivos para cada genótipo com alta probabilidade de associação com as características de interesse econômico particulares a cada um deles (Cardoso-Silva et al., 2014).

Portanto, o objetivo deste estudo é dar continuidade e complementar as análise do transcriptoma gerado a partir do RNA-Seq de folhas de cana-de-açúcar submetidas ao déficit hídrico prolongado, afim de identificar SNPs candidatos dentro de genes relacionados com respostas ao estresse hídrico para que possam ser utilizados como marcadores moleculares na obtenção de cultivares de cana-de-açúcar tolerantes ao estresse hídrico. Para isso, assumimos que polimorfismos específicos dessas regiões de resposta estariam associados com a capacidade das cultivares contrastantes lidarem com o estresse hídrico.

2. REVISÃO DE LITERATURA

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

A cana-de-açúcar cultivada é originária da ilha de Nova Guiné, no centro do Oceano Pacífico, e foi dispersa de forma gradual pela migração humana. No Brasil é considerada planta exótica que foi introduzida pelos colonizadores portugueses na época do descobrimento (Daniels e Roach, 1987).

A cana-de-açúcar pertence ao gênero *Saccharum* L., da tribo Andropogoneae dentro da família das Poaceae, que inclui também as gramíneas tropicais e subtropicais e cereais dos gêneros *Sorghum* (sorgo) e *Zea* (milho). O gênero *Saccharum*, junto com os gêneros *Erianthus* divisão *Ripidium*, *Miscanthus* divisão *Diandra*, *Narenga* e *Sclerostachya*, formam o chamado *Complexo Saccharum*, todos com origem comum, intercruzáveis, altos níveis de ploidia e frequente desbalanço de cromossomos (aneuploidias), dificultando as classificações taxonômicas (Daniels e Roach, 1987; Sreenivasan et al., 1987).

As variedades cultivadas de cana-de-açúcar são híbridos poliploides provenientes do cruzamento de diferentes espécies do gênero *Saccharum*. Os híbridos são resultantes de cruzamentos interespecíficos realizados na primeira metade do século XX entre *S. officinarum* e *S. spontaneum* e têm entre 100 a 130 cromossomos, sendo 80% derivados de *S. officinarum*, 10% de *S. spontaneum* e 10% de recombinantes das duas espécies (Piperidis et al., 2010). O cruzamento interespecífico envolve a combinação, melhoramento do vigor e a resistência a doenças a partir de *S. spontaneum* e alta taxa de acúmulo de sacarose presente em *S. officinarum*, incrementando assim, o conteúdo de sacarose e adquirindo resistência a doenças nos cultivares comerciais (D'hont et al., 1996).

Entretanto, o melhoramento genético da cultura da cana-de-açúcar não é tão simples, pois envolve outras características e técnicas além daquelas que visam melhorar a produtividade de sacarose, por exemplo, produção por unidade de área, resistência ao florescimento e à escassez de água, entre outros. Além disso, o alto nível de ploidia torna o processo de melhoramento ainda mais complexo (Rodrigues et al., 2009; Rodrigues et al., 2011).

Em relação à produção, o crescimento e o desenvolvimento da cultura consiste puramente no desenvolvimento de órgãos vegetativos, colmos e raízes, definidos em quatro estágios distintos (Figura 1): brotação e emergência, perfilhamento, crescimento dos colmos e maturação dos colmos.

O primeiro estágio, brotação e emergência, depende da qualidade da muda, ambiente, época e manejo do plantio. É neste estágio que ocorrem o enraizamento inicial e o aparecimento das primeiras folhas. A fase de perfilhamento é o segundo estágio de desenvolvimento, onde ocorre a emissão de colmos por uma mesma planta, os quais recebem a denominação de perfilhos e que, por sua vez, darão origem às touceiras da cana-de-açúcar. O terceiro estágio corresponde à fase de crescimento dos colmos, caracterizada pelo crescimento e desenvolvimento dos mesmos, que ganham altura e começam a acumular açúcar na base. Essa fase é estimulada por luz, umidade e calor. O último estágio, denominado de maturação, caracteriza-se pelo intenso acúmulo de sacarose nos colmos, determinando a qualidade da matéria-prima dos colmos industrializáveis (Gascho e Shih, 1983).

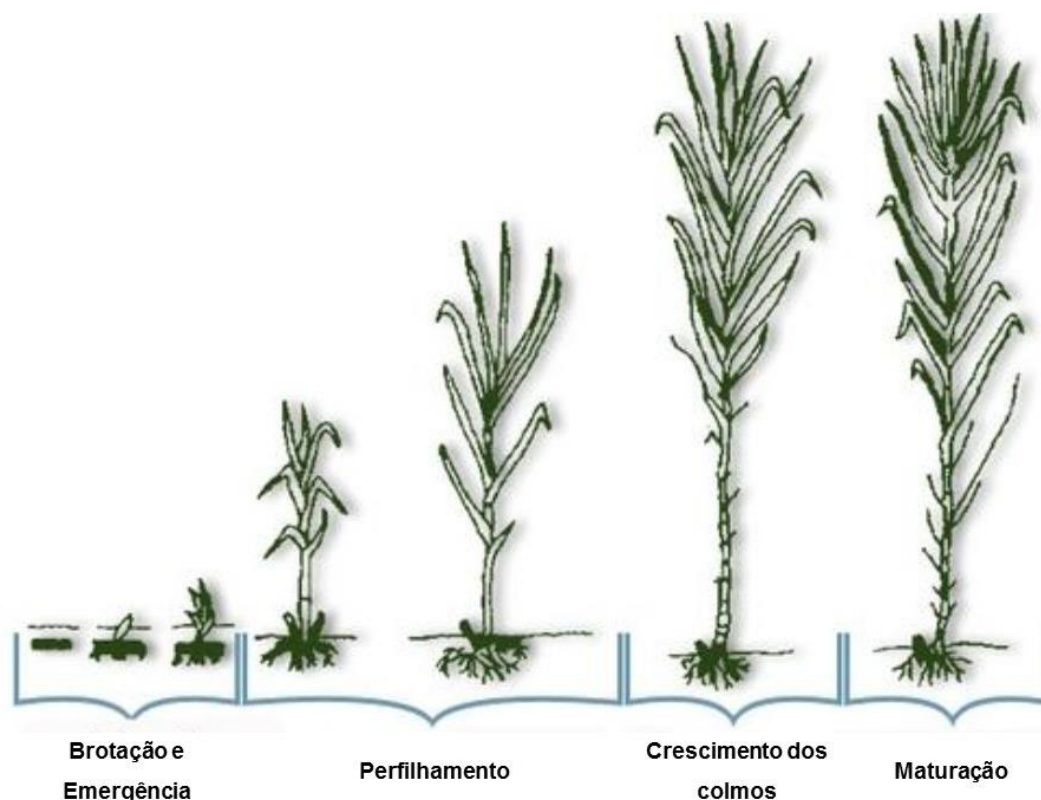


Figura 1. Estágios de desenvolvimento da cana-de-açúcar. Adaptado de Gascho e Shih (1983).

Por se tratar de uma planta de metabolismo C4, a cana-de-açúcar é considerada altamente eficiente na conversão de energia radiante em energia química. A característica de alta produtividade de fotoassimilados em plantas C4 é atribuída à cooperação metabólica de células das folhas fotossinteticamente ativas, que constituem a anatomia do tipo Kranz, na bainha do feixe vascular (Hatch, 1987). Geralmente, plantas de metabolismo C3 conseguem converter menos de 1% da radiação incidente em energia química armazenada através da fotossíntese (McKendry, 2002), enquanto a cana-de-açúcar é capaz de converter aproximadamente 2% da luz solar disponível em carboidratos, onde dois terços destes carboidratos estão na forma de lignocelulose e o restante está na forma de açúcares solúveis, principalmente sacarose (Singh et al., 2008).

O clima tropical do Brasil apresenta altas temperaturas na maior parte do ano, o que favorece o cultivo de plantas de metabolismo C4 como a cana-de-açúcar. Além disso, outros fatores como umidade, luz, solo e disposição de nutrientes contribuem para que o país seja o maior produtor mundial de cana-de-açúcar, com cerca de 620 milhões de toneladas processadas na safra 2018/2019. A Região Centro-Sul (que agrega os estados das regiões Sul, Sudeste e Centro-Oeste) responde por 90% deste volume, enquanto os 10% restantes cabem aos estados da região Norte-Nordeste (Figura 2) (Unica, 2019).

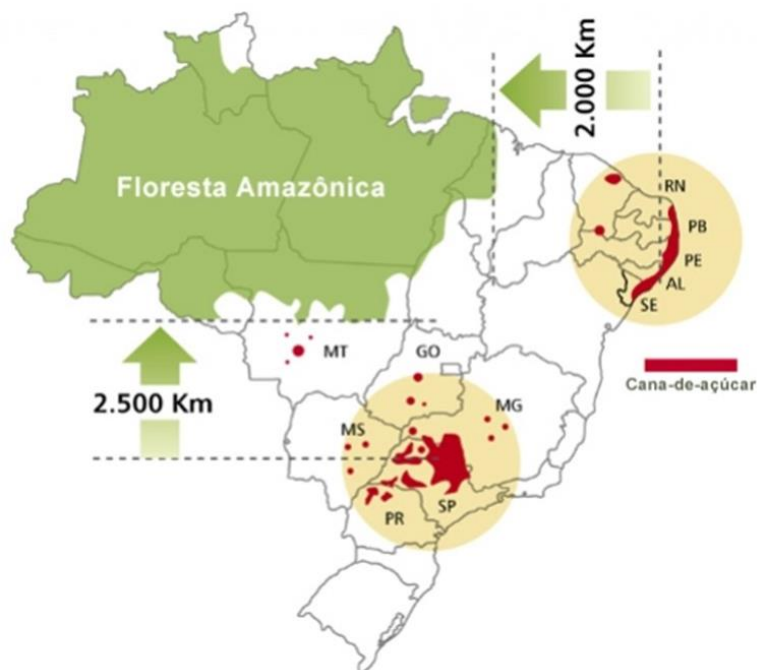


Figura 2. Mapa de produção da cana-de-açúcar no Brasil (Unica, 2019).

As projeções para o setor sucroenergético nos próximos 10 anos é bastante animadora. De acordo com levantamento feito pela Federação das Indústrias do Estado de São Paulo (FIESP), a safra 2028/2029 de cana-de-açúcar deve apresentar aumentos de 36% na produção da planta, 23% na produtividade e 11% na área plantada em relação à safra 2018/2019. Além disso, a frota brasileira de veículos leves terá um aumento de cerca de 10 milhões de unidades, onde 31% desses veículos serão “flex”, aumentando a demanda por etanol hidratado.

Essa perspectiva de aumento de demanda tem levado a produção de cana-de-açúcar para áreas pouco adaptadas, como a região Centro-Oeste do Brasil, onde os meses de abril a novembro são predominantemente secos. Esse período compreende a fase final de crescimento vegetativo, a fase de maturação e o início do período de perfilhamento da cultura. Além disso, as mudanças climáticas têm provocado eventos extremos, ora com anos mais chuvosos, ora com anos mais secos. Dessa forma, torna-se cada vez mais importante o desenvolvimento e a introdução de novas cultivares tolerantes à seca para diminuir os impactos na produtividade dos canaviais.

2.2. Estresse hídrico em plantas

Desde os antigos povos sumérios, o homem tem procurado uma alternativa mais efetiva do aproveitamento da água para superar os efeitos do estresse hídrico às plantas. Esse tipo de estresse é um dos principais gargalos para o crescimento e desenvolvimento de plantas cultivadas em todo o mundo, junto com outros estresses abióticos como a salinidade, temperatura e radiação extremas, contaminação por metais pesados e deficiência ou excesso de nutrientes minerais (Santos e Carlesso, 1998; Ahmad et al., 2016).

A água representa cerca de 90% da massa de matéria verde nos vegetais, constituindo a matriz e o meio onde ocorrem grande parte dos processos bioquímicos necessários para a vida da planta. A disponibilidade de água é o recurso mais abundante que as plantas necessitam para crescer e funcionar e, paralelamente, é também o fator mais limitante, influenciando diretamente na produtividade (Taiz et al., 2017).

De acordo com a União das Indústrias de Cana-de-Açúcar (Unica), somente a

falta de água foi responsável por uma redução de 6,05% na produtividade de cana-de-açúcar na safra 2018/2019. Esse número equivale a 36 milhões de toneladas de cana que deixaram de ser produzidas. Para se ter uma ideia, esse número representa toda a produção do estado do Paraná, o quinto maior produtor de cana-de-açúcar do país.

Ainda com relação à disponibilidade hídrica, tanto a falta como o excesso de água podem causar danos às plantas, sendo o estresse por deficiência o mais recorrente, ocasionando mudanças e respostas em diferentes níveis funcionais do organismo, as quais podem ser reversíveis no início e, mais tarde, tornam-se permanentes caso a falta de água persista (Larcher, 2000; Pincelli, 2010).

A deficiência hídrica pode ser definida como todo o conteúdo de água de um tecido ou célula que está abaixo do conteúdo de água mais alto exibido no estado de maior hidratação (Taiz et al., 2017). Segundo Lawlor e Cornic (2002), as plantas absorvem água através das raízes e a mesma é constantemente perdida para a atmosfera através do processo de evapotranspiração. Assim, quando a perda de água consequente da evapotranspiração é maior que a absorção pelas raízes, a planta entra em déficit hídrico.

As plantas de cana-de-açúcar em condições de deficiência hídrica manifestam inúmeras alterações morfofisiológicas que levam à uma desordem do seu metabolismo, destacando-se: (i) alterações na altura, (ii) número de folhas verdes, (iii) comprimento e largura das folhas, (iv) área foliar e massa foliar específica, (v) densidade estomática, (vi) condutância estomática, (vii) eficiência quântica do fotossistema II, (viii) teor relativo de água, (ix) conteúdo de clorofila e (x) alterações no acúmulo de matéria seca da parte aérea e das raízes. Todo esse desequilíbrio funcional afeta os processos metabólicos da planta, trazendo consequências para a qualidade e a quantidade de matéria produzida (Pimentel, 2004; Taiz et al., 2017; Rodrigues et al., 2018).

Ao longo da sua evolução, as plantas desenvolveram complexos mecanismos para que elas pudessem se adaptar às situações de estresses bióticos e abióticos. Dessa forma, a aptidão para resistir ao déficit hídrico é um importante determinante na distribuição natural das plantas e na produtividade das culturas agrícolas, tornando a disponibilidade de água um fator seletivo na evolução das espécies (Lawlor e Cornic

2002; Bartels e Sunkar, 2005).

As plantas que realizam fotossíntese através do metabolismo C4, como a cana-de-açúcar e o milho, são exemplos desses mecanismos adaptados desenvolvidos para situações de estresse. A presença da bainha vascular e seus mecanismos bioquímicos acoplados à via C3 criaram, durante a evolução, uma espécie de “bomba” que torna o sistema fotossintético mais eficiente em certas situações, praticamente eliminando a fotorrespiração, uma via metabólica dispendiosa que ocorre quando a enzima RUBISCO do Ciclo de Calvin atua sobre o oxigênio em vez do dióxido de carbono. Essas plantas conseguem ter maior aproveitamento de CO₂, garantindo maior controle da abertura dos estômatos e evitando a perda de água pelos mesmos (Edwards e Huber, 1981; D’Hont et al., 2008).

Em estudo sobre os efeitos da seca em genótipos de milho, Atteya (2003) constatou uma redução na condutância estomática logo no início do estresse hídrico, sem que houvesse qualquer alteração no potencial hídrico de água na folha. De acordo com Yordanov et al. (2003), essa resposta no comportamento dos estômatos em plantas que foram submetidas ao estresse hídrico já era esperada, sendo essa uma das primeiras estratégias da planta para impedir a desidratação excessiva das folhas.

Em casos de estresse hídrico severo em plantas de metabolismo C4, a taxa fotossintética é reduzida, principalmente, devido a restrições consideradas não estomáticas ou metabólicas, como a inibição da divisão celular e a síntese de proteínas (Galmés et al., 2011; Taiz et al., 2017).

Quando a disponibilidade de água à planta é suspensa abruptamente, os mecanismos morfofisiológicos são severamente afetados. No entanto, quando o déficit hídrico ocorre gradualmente, a planta passa por adaptações para lidar com o estresse, como redução no crescimento, diminuição no tamanho das folhas e senescência foliar por exemplo. Essas adaptações ocorrem, principalmente, quando a planta se encontra no início do ciclo, diminuindo a perda de água para a atmosfera (Chaves, 1991; Chaves et al., 2009).

A identificação das características morfofisiológicas e dos mecanismos moleculares associados à tolerância à seca são abordagens úteis em estudos de tolerância e resposta ao estresse, tornando-se indiscutível a importância desse tipo

de estudo para o desenvolvimento de plantas tolerantes ao estresse hídrico e que apresentem maior eficiência no uso da água (Osakabe et al., 2014).

2.3. Mecanismos de resistência ao estresse

As plantas apresentam diferentes mecanismos de resistência ao estresse hídrico que podem ser divididos em escape, retardo e tolerância. No primeiro, as plantas manifestam rápido desenvolvimento fenológico e alto grau de plasticidade, completando seu ciclo de vida antes que a falta de água se torne severa o bastante para provocar danos fisiológicos. O mecanismo de retardo refere-se à manutenção do turgor e do volume celular, tanto pelo desenvolvimento de um sistema radicular abundante para absorção de água, quanto pela redução da perda por transpiração por meio do fechamento estomático ou por vias não estomáticas como a cutícula. Já a tolerância permite que a planta mantenha seu metabolismo funcionando normalmente, mesmo que o potencial hídrico dos tecidos esteja reduzido, devido principalmente ao acúmulo de solutos compatíveis ou osmólitos, proteínas osmoprotetoras e à capacidade antioxidante (Verslues et al., 2006; Taiz et al., 2017).

A tolerância à seca é uma característica quantitativa que envolve uma complexa rede de respostas em nível morfológico, através da redução da área foliar e aumento do sistema radicular, em nível fisiológico, por meio do fechamento estomático, ajuste osmótico e maior eficiência no uso da água, e em nível molecular, pela expressão diferencial de genes relacionados à resposta ao estresse (Bray, 1993; Taiz et al., 2017).

As plantas podem responder ao estresse hídrico de maneiras diferentes dependendo da duração e da severidade da exposição, da idade e do estágio fenológico em que se encontram (Chaves et al., 2003). Assim, as plantas podem apresentar mecanismos de tolerância ou resistência e suscetibilidade ao estresse hídrico, podendo chegar à morte, dependendo da severidade do estresse a que as mesmas são submetidas (Chaves et al., 2003; Cambraia, 2005).

O estresse é percebido pelas plantas através de receptores presentes na membrana das células vegetais, como íons de Ca^{2+} , sensores de Ca^{2+} , proteínas quinases dependentes de Ca^{2+} (CDPKs), espécies reativas de oxigênio (EROs),

enzimas que clivam fosfolipídios, entre outros. Essas proteínas interagem com seus respectivos parceiros de interação, muitas vezes iniciando uma cascata de fosforilação e têm como alvo os principais genes responsivos ao estresse ou os fatores de transcrição (FTs) que controlam esses genes, incluindo as famílias de FTs CBF/DREB, ABF/ABAE, HSF, bZIP e MYC/MYB. Os fatores de transcrição, por sua vez, ativam genes cujos produtos atuam diretamente na proteção de membranas e proteínas e no restabelecimento da homeostasia celular, como as proteínas de choque térmico (HSPs), chaperonas e as proteínas LEA, aquaporinas e os transportadores de íons por exemplo (Wang et al., 2003). Essa complexa rede de respostas está exemplificada na Figura 3.

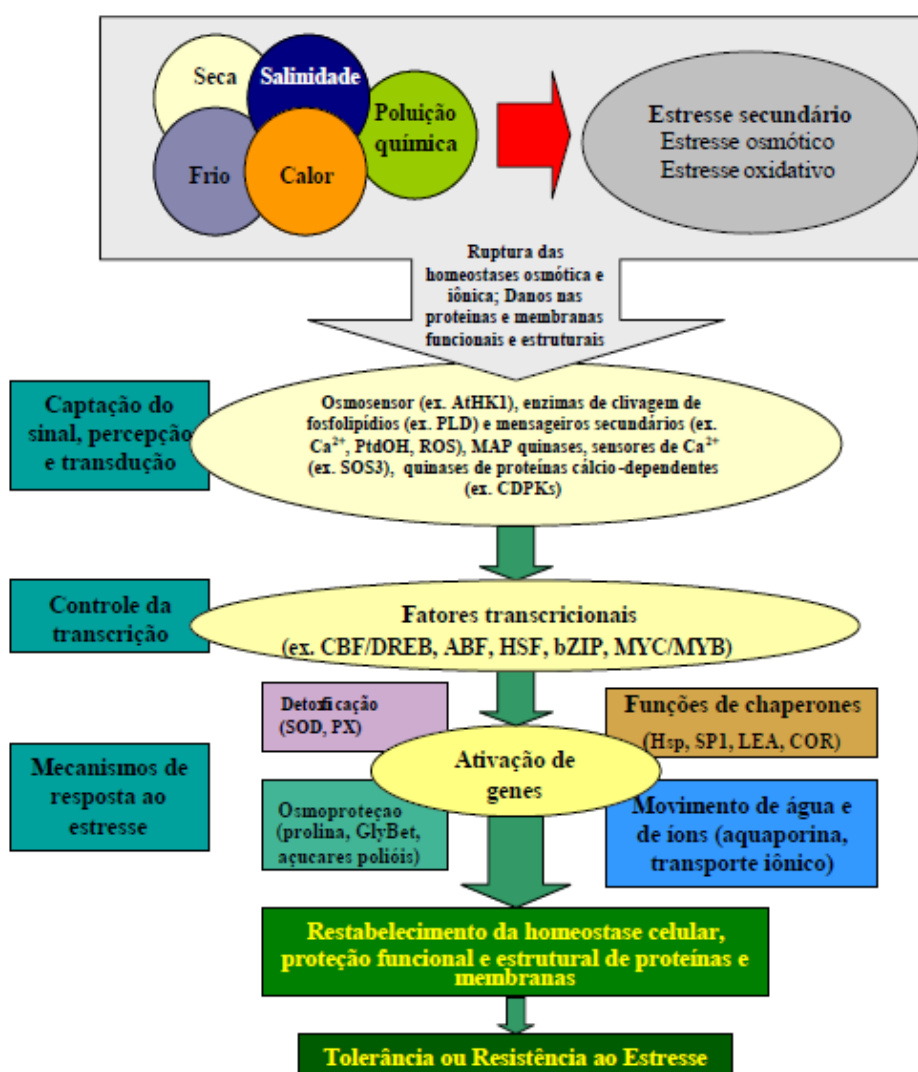


Figura 3. Representação esquemática dos mecanismos de defesa desencadeados pelas plantas em resposta à estresses abióticos. Adaptado de Wang et al. (2003).

Segundo Mahajan e Tuteja (2005), as alterações na expressão gênica provocadas pela falta de água podem induzir a síntese de fitormônios como ácido abscísico (ABA), ácido salicílico, ácido jasmônico e etileno. Essas moléculas podem amplificar a sinalização do estresse e desencadear uma nova rota de transdução de sinal, auxiliando na resposta de defesa da planta. O autor também afirma que pode haver participação indireta de moléculas acessórias na sinalização, incluindo modificadores de proteínas como enzimas de glicosilação, metilação e ubiquitinação.

A tolerância das plantas ao estresse hídrico não é uma característica simples, mas uma característica onde mecanismos trabalham isoladamente ou em conjunto para evitar ou tolerar períodos de seca. Uma resposta fisiológica específica ao déficit hídrico representa a combinação de eventos moleculares prévios, que foram ativados pela percepção do sinal de estresse (Nepomuceno et al., 2001). Na maioria das vezes, as mesmas vias de sinalização são ativadas por diferentes tipos de estresse, resultando em respostas parecidas. Entretanto, o mecanismo por trás das respostas das plantas aos estresses abióticos é bastante complexo e envolve a participação de inúmeros genes (Ferro, 2008).

Considerando que a tolerância a seca é uma herança poligênica, torna-se difícil a obtenção de cultivares tolerantes por meio do melhoramento genético clássico. Entretanto, as ferramentas biotecnológicas e moleculares disponíveis, como análise de transcriptoma em larga escala por exemplo, permitem identificar e isolar genes envolvidos nas respostas ao estresse hídrico, auxiliando na obtenção de novas variedades resistentes e tolerantes de cana-de-açúcar.

2.4. Estudo do transcriptoma via RNA-Seq

A descoberta da estrutura do DNA por James Watson e Francis Crick em 1953 foi fundamental para entender como o DNA era capaz de codificar as proteínas responsáveis por todos os processos e pelo fenótipo de todos os seres vivos (Faleiro e Andrade, 2011). Quase 30 anos depois, Frederick Sanger e colaboradores publicaram o primeiro genoma completo de um bacteriófago, abrindo caminho para o sequenciamento de vários outros genomas completos de organismos procariotos e eucariotos (Moreira e Varani, 2015).

Assim surgiu a genômica, ciência que estuda os genomas e que pode ser dividida em genômica estrutural e genômica funcional. Enquanto a genômica estrutural estuda a estrutura dos genomas, a genômica funcional descreve a função específica de genes e proteínas, buscando relacionar as etapas do fluxo da informação biológica, além de relacionar a regulação da expressão gênica com mecanismos bioquímicos e fisiológicos (Fragoso et al., 2011).

Em um mesmo indivíduo, a maioria das células apresentam o mesmo genoma estrutural, ou seja, possuem todas as informações genéticas necessárias para gerar um organismo inteiro. Contudo, ao se especializar, cada célula passa a expressar somente uma parte dessas informações (Costa, 2011). Para isso, a mensagem contida no DNA é convertida em uma função celular através da transcrição dos genes em uma molécula de RNA, que posteriormente poderá ser traduzido em proteínas (Nelson e Cox, 2014).

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 et al., 2006).

Em estudos de genômica funcional são obtidas as informações do que está sendo expresso nas células e tecidos (RNAs e proteínas) do organismo. Assim, as abordagens mais utilizadas nos estudos de genômica funcional têm sido a análise da expressão gênica (transcriptoma) e do padrão proteico (proteoma) codificado pelo genoma estrutural numa determinada condição ambiental (Costa, 2011).

O transcriptoma representa o conjunto completo de transcritos (RNAs mensageiros (RNAm), RNAs ribossômicos (RNAr), RNAs transportadores (RNAt) e os microRNAs) de um dado organismo, órgão, tecido ou linhagem celular, ou seja, ele é o reflexo direto da expressão dos genes (Nelson e Cox, 2014). Através do estudo do transcriptoma, é possível determinar quando e onde os genes são ativados ou desativados em diferentes tipos de células e tecidos, permitindo saber a quantidade de atividade gênica ou expressão existe em uma célula por meio da quantificação do número de transcritos (Wang et al., 2009).

A análise do transcriptoma da cana-de-açúcar permite associar o perfil da

expressão de genes envolvidos com caracteres de interesse econômico, como resistência ao estresse hídrico por exemplo. Diante disso, diversos grupos de pesquisa têm se dedicado a esclarecer os mecanismos que a cana-de-açúcar utiliza para se adaptar ao déficit hídrico, através da análise do transcriptoma das plantas submetidas a esse estresse (Rodrigues et al., 2011; Ferreira et al., 2012; Gentile et al., 2013; Cardoso-silva et al., 2014; Vargas et al., 2014; Vantini et al., 2015; Belesini et al., 2017; Telles et al., 2019).

Diferentes técnicas podem ser utilizadas para o estudo do transcriptoma, com destaque para as tecnologias de nova geração (NGS) que permitem o sequenciamento em larga escala, gerando informação sobre milhões de pares de bases em uma única corrida. Sequenciadores de nova geração e suas plataformas possibilitaram um novo método de sequenciamento de bibliotecas de DNA complementar (cDNA), denominado RNA-seq (Wang et al., 2009), a qual permite o mapeamento e quantificação de transcriptomas, e que vem substituindo rapidamente as antigas abordagens no estudo de expressão gênica (Wang et al., 2009; Fu et al., 2009; Van Vliet, 2010).

Marguerat e Bähler (2010) apontam algumas vantagens da tecnologia de RNA-Seq, como (i) maior precisão na quantificação da expressão dos genes; (ii) a mesma base é sequenciada múltiplas vezes, possibilitando a identificação de SNPs, os quais também podem ser utilizados para estimar a expressão alelo-específica; (iii) é possível identificar “splicing” alternativo ou sequências rearranjadas quando se tem um genoma de referência; e (iv) sua elevada sensibilidade permite a detecção de transcritos mesmo quando os níveis de expressão são baixos. Além disso, essa tecnologia permite o estudo de organismos que não possuem genoma de referência, uma vez que é baseada em sequenciamento de cDNA em larga escala.

O sequenciamento via RNA-Seq resume-se, basicamente, em extração de RNA e construção das bibliotecas de cDNA, amplificação das amostras, sequenciamento e análise dos dados (Wang et al., 2009). Todos os detalhes do processo estão descritos na Figura 4. É importante ressaltar que os protocolos de sequenciamento de RNA variam de acordo com a estratégia e a plataforma de sequenciamento utilizadas no experimento. Neste estudo, foram utilizadas as plataformas Illumina/Solexa para o sequenciamento dos fragmentos no protocolo de

RNA-Seq. Essa tecnologia possibilita que uma população de RNAm seja convertida em uma biblioteca de cDNA fragmentada ao acaso, com adaptadores ligados em uma única extremidade (sequenciamento “single-read”) ou em ambas as extremidades (sequenciamento “paired-end”) (Jain, 2012). O sequenciamento “paired-end” foi utilizado nesse trabalho porque facilita a detecção de rearranjos genômicos, elementos de sequências repetidas, assim como fusões de genes e novos transcritos. As leituras geradas a partir do sequenciamento “paired-end” têm maior probabilidade de se alinharem a uma referência, o que melhora a qualidade do conjunto de dados (Wang et al., 2009).

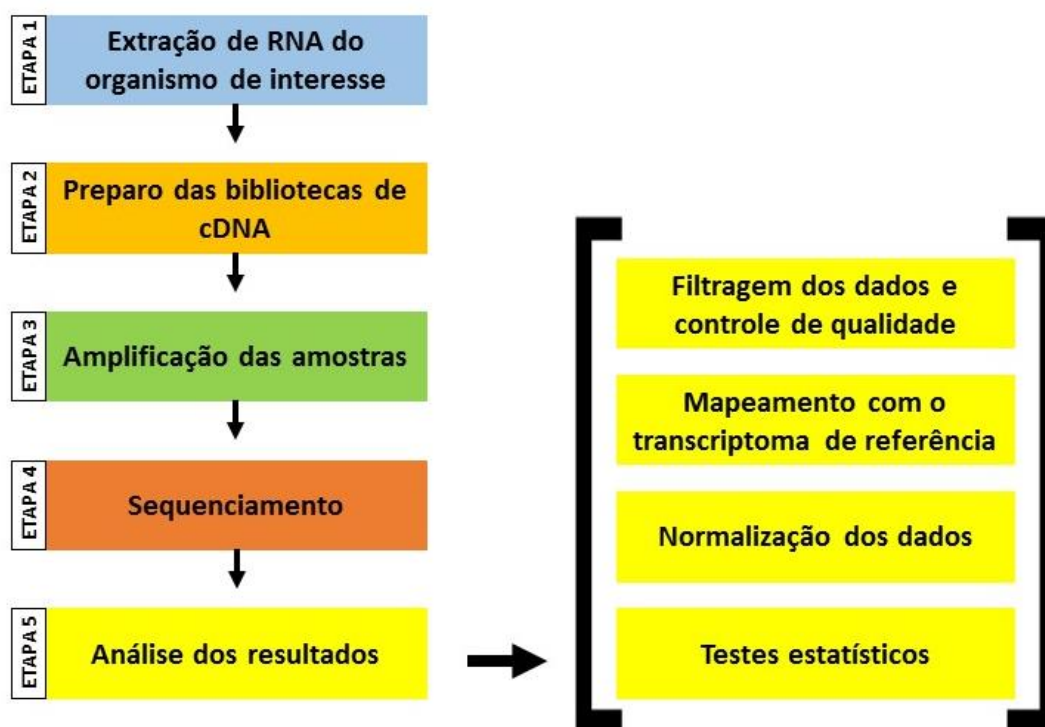


Figura 4. Esquema geral de um experimento utilizando a técnica de RNA-Seq na plataforma Illumina/Solexa. A etapa 1 engloba a preparação da amostra a partir do RNA extraído do organismo de estudo. A etapa 2 consiste na construção de bibliotecas de cDNA a partir das amostras de RNAm purificado. Na etapa 3 ocorre a clonagem “in vitro” para amplificação das amostras através da técnica de PCR. A etapa 4 é o sequenciamento das bibliotecas de cada amostra na plataforma Illumina/Solexa. Por último, a etapa 5 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. Adaptado de Jain (2012).

Para tornar as análises possíveis, a montagem de um transcriptoma referência e criação de um índice gênico bem anotado é fundamental. A primeira iniciativa de criação desse índice gênico para cana-de-açúcar (*Saccharum officinarum*) surgiu no projeto SUCEST, que gerou 238.000 etiquetas de sequências expressas (ESTs) a partir de diversos tecidos de variedades brasileiras, a qual resultou em 43.141 transcritos putativos únicos de 33.000 genes (Vettore et al., 2001; 2003).

Em seguida, surgiram outras iniciativas como o Sugarcane Gene Index (SGI) ou *S. officinarum* Gene Index (SoGI), desenvolvido pelo Institute for Genomic Research (TIGR) e Dana-Farber Cancer Institute (DFCI), a partir de um conjunto de 282.683 ESTs montadas em 42.377 sequências consensos, ou transcritos únicos (Devarumath et al., 2013).

Após a evolução dos sequenciadores, que passaram a ter maior capacidade de rendimento, a quantidade de dados de fragmentos de sequências de transcritos gênicos aumentou de forma exponencial para diversas espécies, inclusive para a cana-de-açúcar. Entretanto, essas bases de dados, tais como o SoGI e SGI, deixaram de ser atualizadas, tornando as sequências existentes defasadas para utilização como referência nas análises de transcriptomas de cana-de-açúcar, em especial nos estudos de expressão gênica.

Além disso, surgiram novas estratégias para montagem de transcriptomas com base em dados de RNA-Seq, as quais levam em consideração as características das plantas não-modelos, tais como a cana-de-açúcar. O RNA-seq possibilita ainda a identificação de SNPs em regiões transcritas, os quais podem ser utilizados na detecção de expressão alelo-específica e também como marcadores moleculares (Cardoso-Silva et al., 2014).

Através dessa tecnologia é possível identificar SNPs exclusivos para cada genótipo com alta probabilidade de associação com as características de interesse econômico particulares a cada um deles (Cardoso-Silva et al., 2014). Tais estratégias têm como objetivo a busca de SNPs em regiões codificadoras, por meio do sequenciamento em larga escala do transcriptoma de tecidos alvo de cultivares ou variedades contrastantes para características de interesse, como tolerância à seca por exemplo (Giachetto e Higa, 2014).

2.5. Marcadores Moleculares do tipo SNP (*Single Nucleotide Polymorphisms*)

Marcadores moleculares são sequências de DNA que revelam polimorfismos em regiões codificantes ou não codificantes de um gene, considerando uma população de indivíduos. Estas sequências estão distribuídas aleatoriamente ao longo do genoma e permitem acessar diretamente o genótipo de interesse ao invés do fenótipo, evitando qualquer tipo de influência ambiental, como ocorre na maioria dos programas de melhoramento clássico (Ferreira e Grattapaglia, 1998; Kumar, 1999).

Essa ferramenta tem sido amplamente utilizada na cultura da cana-de-açúcar em análises de divergência genética (Santos et al., 2012), análises filogenéticas (Prasitsom et al., 2019), mapeamento genético (Marconi et al., 2011), além do melhoramento genético visando resistência a estresses bióticos e abióticos (Markad et al., 2014; McCord et al., 2019).

Todos esses estudos têm como objetivo comum a busca por variações em sequências gênicas que possam estar relacionadas com um fenótipo observado, além de minimizar o tempo necessário para a obtenção de uma nova cultivar através do melhoramento clássico, principalmente para culturas perenes. Portanto, o estudo dos marcadores moleculares possibilita a identificação precoce e precisa de indivíduos com uma melhor combinação de alelos favoráveis, melhorando significativamente a análise genética (Gibson, 2006; Henry, 2012).

Os marcadores de DNA podem ser divididos em três categorias principais: os baseados em hibridização, os baseados em PCR (Reação em cadeia da Polimerase) e por último, os marcadores baseados em sequenciamento. Os marcadores também podem ser classificados de acordo com o tipo de herança alélica em dominantes e codominantes. Os marcadores codominantes possibilitam diferenciar indivíduos homozigotos e heterozigotos, o que não é possível quando se utiliza marcadores dominantes, onde apenas é possível identificar a presença ou ausência de um determinado alelo (Turchetto-Zolet et al., 2017).

Os principais exemplos de marcadores dominantes são o AFLP (Polimorfismo de Comprimento de Fragmentos Amplificados - “Amplified Fragment Length Polymorphism”) (Zabeau e Vos, 1993) e o RAPD (Polimorfismo do DNA Amplificado ao Acaso - “Random Amplified Polymorphism DNA”) (Williams et al., 1990), ambos

baseados em PCR. Entre os marcadores de expressão codominante, destacam-se o RFLP (Polimorfismo no Comprimento dos Fragmentos de Restrição - “Restriction Fragment Length Polymorphism”), baseado em hibridação (Botstein et al., 1980), os microssatélites ou SSR (Sequências Simples Repetidas - “Simple Sequence Repeats”), baseado em PCR (Hamada et al., 1982), e os SNPs (Polimorfismo de Único Nucleotídeo - “Single Nucleotide Polymorphism”), baseado em sequenciamento (Gupta et al., 2001).

Segundo Borém e Caixeta (2009), a escolha de qual marcador utilizar deve levar em conta a tecnologia utilizada para revelar a variabilidade em nível do DNA, a capacidade de gerar diferenças entre indivíduos (polimorfismo), o custo, a facilidade de uso e a reprodutibilidade.

Os marcadores do tipo SNP apresentam algumas vantagens quando comparados aos marcadores convencionais anteriormente citados, como os RFLPs, RAPDs, AFLPs e SSRs. Esses marcadores convencionais são desenvolvidos, principalmente, nas regiões não codificantes dos genes de interesse, ao contrário dos SNPs, que são geralmente desenhados com base nos polimorfismos dentro das regiões transcritas de genes e conseqüentemente correlacionados com a função desses genes (Andersen e Lübberstedt, 2003). Além disso, os SNPs são passíveis de automação em larga escala, apresentam alta adaptabilidade a diversas tecnologias e a necessidade de manipulação laboratorial é menor (Hara et al., 2010; Weller et al., 2010).

Os SNPs constituem o tipo mais básico de variação genética, revelando polimorfismos específicos que ocorrem em uma única posição no genoma, em um único nucleotídeo (Gupta et al., 2001). Alguns autores também consideram “Indels” (inserções ou deleções de nucleotídeos) como SNPs, embora muito provavelmente eles ocorram por mecanismos diferentes (Kahl et al., 2005). Esses polimorfismos podem ser detectados com a análise de ESTs disponíveis (análises “in silico”), nas sequências de genes candidatos (análise “in vivo”) ou nas sequências de DNA genômico (DNAG).

Geralmente, esse tipo de marcador é bialélico, ou seja, uma base nucleotídica (A, C, T ou G) substitui a posição de uma das outras três, sendo, portanto, amplamente distribuídos e conservados ao longo do genoma. Além disso, os SNPs também são

classificados de acordo com a variação da base em transições e transversões. As primeiras representam trocas entre bases púricas (A/G) ou entre bases pirimídicas (C/T), enquanto as transversões consistem em trocas de bases púricas por bases pirimídicas e vice-versa. Eles também podem ocorrer tanto em sequências codificantes como não codificantes (Rafalsky, 2002; Kahl et al, 2005).

Os SNPs são marcadores altamente polimórficos, e podem ser utilizados em diversos estudos genéticos, como construção de mapas genéticos, mapeamento de locus de características quantitativas (QTL), caracterização de indivíduos de uma população, seleção assistida por marcadores (SAM), análises de diversidade, análises filogenéticas e estudos de tolerância a estresses bióticos e abióticos (Garcia et al., 2013).

Os SNPs podem ser classificados em SNPs codificantes ou SNPs não codificantes, de acordo com a região em que se encontram. Os SNPs não codificantes podem estar localizados dentro de íntrons, ou também podem ocorrer em regiões regulatórias ou promotoras do genoma, denominados então de SNPs intrônicos, SNPs reguladores ou SNPs promotores, respectivamente. SNPs em regiões UTRs ou intrônicas podem causar alterações na tradução das proteínas ou na produção de transcritos variantes por ocorrência ou não de “splicing” (encadeamento alternativo de éxons, ou “alternative splicing”), levando a sequências de proteínas mais longas ou mais curtas, respectivamente. Já os SNPs promotores podem afetar a atividade e regulação do gene dirigido por esse promotor, por exemplo, impedindo a ligação de um fator de transcrição na sua sequência de reconhecimento, alterando a expressão desse gene (Nogales e DeDiego, 2019).

Os SNPs presentes nas regiões codificantes dos éxons podem causar algum impacto na função da proteína codificada quando a variação implica na troca para um códon que codifica um aminoácido distinto, sendo nesse caso chamado de SNP não-sinônimo. Por outro lado, a presença de um SNP nessa mesma região pode ser totalmente neutra, ou seja, não altera a composição de aminoácidos do domínio ou da proteína codificada e, por isso não apresenta um efeito sobre a sua função, e nestes casos, o SNP é chamado de SNP sinônimo. Apesar dos SNPs resultantes de mutações sinônimas não modificarem a composição de aminoácidos, eles podem acometer o dobramento de proteínas, afetando o posicionamento de seus domínios

de ligação, por exemplo (Kahl et al., 2005).

De acordo com Cardoso-Silva et al. (2014), os estudos envolvendo SNPs em cana-de-açúcar começaram a partir da disponibilização de coleções de ESTs nos bancos de dados. Margulies et al. (2005) foram os pioneiros, identificando SNPs em mais de 200 genes nos genitores de uma população de mapeamento. Cordeiro et al. (2006) mostraram que os SNPs ocorrem com alta frequência no genoma da cana-de-açúcar, identificando os marcadores em vários genes diferencialmente expressos de colmos maduros e imaturos. Garcia et al. (2013) identificaram SNPs a partir de sequências expressas, mostrando uma nova modelagem genética estatística e melhor compreensão dos aspectos biológicos dos processos evolutivos e de domesticação.

Neste trabalho realizamos a identificação dos SNPs candidatos através do “software” FreeBayes (Garrison e Marth, 2012) para identificação de sítios variantes a partir de dados de sequenciamento. Esse programa utiliza o método Bayesiano para encontrar polimorfismos menores que o comprimento do alinhamento de uma sequência de leitura curta, como SNPs, “Indels” e eventos similares. Seu funcionamento é baseado em haplótipos, detectando os sítios variantes de acordo com sequências de leituras alinhadas a um alvo particular. O resultado é um conjunto de posições em que se encontram prováveis polimorfismos em formato VCF (“variant call file format”), um arquivo de texto utilizado em bioinformática para armazenar variações de sequências genômicas (Danecek et al., 2011).

Para a anotação funcional desses polimorfismos utilizamos o “software” SnpEff que, segundo Cingolani et al. (2012), confronta a posição em que cada variação está situada com anotações pré-existentes de uma sequência de referência. Dentre as atribuições realizadas por esse programa destaca-se a predição de efeitos com os seguintes graus de impacto: alto, moderado, baixo e modificador.

Cingolani et al. (2012) define como de alto efeito as mutações que provocam a excisão do transcrito, perda ou ganho de códon de terminação, ganho de códon de iniciação, alteração em sítio doador ou receptor de “splicing”, alteração para um códon que traduz um aminoácido raro, perda de éxon e variação no quadro de leitura ou no número de cromossomos. Moderados são os efeitos que produzem o truncamento de UTR com consequente perda de éxon, variação de baixo impacto na sequência codante, deleções ou inserções sem maiores consequências, simples alteração de

aminoácido (mutação de sentido trocado), excisão de região regulatória, perda de região de “splicing” e excisão de regiões de ligação de fatores de transcrição.

As alterações de baixo efeito são descritas para os casos de mutação silenciosa, mutação que retém o códon de iniciação ou de terminação, modificações com baixo efeito na região de “splicing” e criação prematura de 5'-UTR. Por fim, as mutações de efeito modificador abrangem um amplo espectro de condições, incluindo variações genéricas em regiões não traduzidas que não alteram propriamente a proteína, como alterações em íntrons e na transcrição de micro-RNA.

Por último, é importante ressaltar a necessidade de validação desses SNPs. Segundo Brookes (1999), para que um SNP seja considerado verdadeiro é necessário que ele ocorra em no mínimo 1% de uma determinada população. Portanto, é preciso determinar a frequência do SNP em uma população para certificar sua natureza polimórfica (Freire, 2010). Neste trabalho foi realizado um levantamento dos SNPs presentes nas cultivares SP81-3250 e RB85-5453, por isso todos os SNPs citados aqui são considerados candidatos ou putativos.

2.6. Estudos fisiológicos, genéticos e moleculares em cana-de-açúcar submetida a prolongada limitação hídrica

A grande maioria dos estudos envolvendo estresse hídrico severo em cana-de-açúcar em casa de vegetação são, na maioria das vezes, aplicados em um curto período de tempo para identificar e propor cultivares tolerantes (Rodrigues et al., 2009; Gupta et al., 2010; Rodrigues et al., 2011). Assim, nosso grupo de pesquisa desenvolveu um projeto com auxílio da FAPESP (aprovado em 2013, Processo No. 2013/11.617-9), com o objetivo de avaliar o comportamento fisiológico de duas cultivares de cana-de-açúcar com diferentes padrões de tolerância ao déficit hídrico ao longo de um período de 90 dias de estresse, no período conhecido como fase de formação da cana-de-açúcar, compreendendo o período mais crítico por demanda de água.

Foram utilizadas duas cultivares, SP81-3250 e RB85-5453, como padrão de tolerância e suscetibilidade ao déficit hídrico, respectivamente (Pincelli e Silva, 2012). Ambas foram submetidas a três potenciais hídricos do solo: (1) controle (sem déficit

hídrico), (2) déficit hídrico moderado e (3) déficit hídrico severo a partir de 60 dias após o plantio e, em seguida, foram amostradas aos 30, 60 e 90 dias após a aplicação dos tratamentos e avaliadas em relação a parâmetros fisiológicos (número de perfilhos, área foliar, potencial hídrico foliar, índice de cor verde da folha, fotossíntese e concentração interna de CO₂).

É conveniente ressaltar que todas as variáveis fisiológicas analisadas foram eficientes em distinguir as cultivares SP81-3250 e RB85-5453 quanto à tolerância e a sensibilidade à deficiência hídrica, respectivamente, visto que a cultivar tolerante apresentou menores reduções dos parâmetros avaliados que a cultivar sensível (Telles et al., 2019).

O sequenciamento das amostras através da tecnologia de RNA-Seq retornou um total de 177.509 transcritos para a cultivar mais tolerante (SP81-3250) e 185.153 transcritos para a cultivar sensível (RB85-5453). O alinhamento desses transcritos com sequências de outras espécies disponíveis em bancos públicos permitiu identificar o conjunto de genes de cana-de-açúcar compartilhados com outras espécies e identificar novos transcritos ainda não catalogados.

Através desses dados, Belesini et al. (2017) e Konrad (2019) realizaram uma investigação, a nível molecular, de quais transcritos poderiam estar envolvidos na resposta das plantas ao déficit hídrico prolongado, utilizando dados de expressão diferencial. Embora os transcriptomas das cultivares SP81-3250 e RB85-5453 em diferentes condições de déficit hídrico e em diferentes épocas tenham gerado uma grande quantidade de dados, ainda há muito a ser explorado neste acervo de transcritos, como os polimorfismos que ocorrem na região codificadora dos genes de resposta ao estresse hídrico por exemplo, que foram abordados neste estudo e poderão, futuramente, auxiliar no desenvolvimento de cultivares de cana-de-açúcar tolerantes ao estresse hídrico.

3. MATERIAL E MÉTODOS

3.1. Experimento em casa de vegetação, delineamento experimental e material vegetal

O experimento de tolerância ao estresse hídrico teve início em 2013 com o projeto multidisciplinar intitulado “Mecanismos fisiológicos em cultivares de cana-de-açúcar submetidas à prolongada limitação hídrica”, o qual foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - Processo No. 2013/11.617-9) e 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, Dra. Thais Ramos da Silva, MSc. Aline Andruccioli Belesini e MSc. Bruna Robiati Telles.

Aqui descrevemos brevemente o material vegetal utilizado, o delineamento experimental, obtenção das amostras de RNA e a técnica de RNA-Seq. Todos os demais detalhes estão descritos em estudos anteriormente publicados pelo nosso grupo de pesquisa (Belesini et al., 2017; Telles et al., 2019).

O experimento em casa de vegetação foi conduzido com delineamento experimental em blocos casualizados, com esquema fatorial $2 \times 3 \times 3$ (duas cultivares (C) \times três tensões hídricas do solo (T) \times três épocas amostrais), com três repetições cada, totalizando 18 unidades experimentais e 54 amostras. Foram utilizadas duas cultivares comerciais de cana-de-açúcar consideradas potenciais em produtividade segundo Pincelli e Silva (2012): a cultivar SP81-3250, tolerante à seca, e a cultivar RB85-5453, sensível à deficiência hídrica.

A cultivar SP81-3250 apresenta alta produtividade e alto teor de BRIX, enquanto a cultivar RB85-5453 apresenta características mais rústicas, como resistência ao tombamento e a algumas das doenças mais problemáticas na cultura da cana-de-açúcar, inclusive ferrugem, mosaico e carvão. Marin (2019) realizou um levantamento das características agrônômicas das variedades de cana-de-açúcar mais plantadas na Região Centro-Sul do Brasil, inclusive as utilizadas neste estudo (SP81-3850 e RB85-5453), cujas particularidades estão descritas na Tabela 1.

Tabela 1. Características agrônômicas das cultivares utilizadas neste estudo. Adaptado de Marin (2019).

| Descritivo | SP81-3850 | RB85-5453 |
|----------------------------|-----------------------|--------------------|
| Destaque | Rica e produtiva | Rica e ereta |
| Regime hídrico | Tolerante ao estresse | Exigente em água |
| Exigência de solos | ND | Exigente |
| Maturação | Média | Precoce |
| Rendimento de transporte | Bom | Bom |
| Colheita mecânica | Boa | Boa |
| Brotação | Sem restrição | ND |
| Brotação de soca com palha | Excelente | Excelente |
| Fechamento de entrelinhas | Bom | ND |
| Nematoides | Resposta instável | ND |
| Florescimento | Regularmente | Todos os anos |
| Maturadores | Resposta instável | Resposta excelente |

ND: Não descrito

As mudas utilizadas no estudo foram obtidas a partir de toletes de cana-de-açúcar plantados em recipientes plásticos e cultivados até os 60 dias de idade sem qualquer restrição hídrica. Após esse período iniciou-se a aplicação dos tratamentos com três potenciais hídricos de solo distintos, monitorado por tensiometria digital (Figura 5): sem déficit hídrico (-0,010 a -0,015 MPa); potencial hídrico moderado (-0,050 a -0,055 MPa) e potencial hídrico severo (-0,075 a -0,080 MPa).



Figura 5. Condução do experimento em casa de vegetação após aplicação das três distintas tensões hídricas (sem restrição hídrica, restrição hídrica moderada e restrição hídrica severa) – A) Cinco dias após aplicação dos tratamentos, B) Sessenta dias após aplicação dos tratamentos e C) Noventa dias após a aplicação dos tratamentos. Imagem retirada de Telles (2016).

A coleta das folhas +1 para obtenção do RNA total ocorreu aos 30, 60 e 90 dias após o início dos tratamentos (DAT), constituindo-se das idades de 90, 120 e 150 dias após o plantio (DAP).

3.2. Extração, quantificação e análise da integridade do RNA total

Após as coletas das folhas +1 realizadas aos 30, 60 e 90 DAT, foi realizada a extração do RNA total de todas as 54 amostras utilizando-se o kit comercial Purelink RNA Mini Kit da Ambion®. Em seguida, o RNA total extraído foi avaliado quanto à quantidade, qualidade e integridade através dos equipamentos Qubit® 2.0 (Invitrogen), NanoDrop ND-1000 (Thermo Fisher Scientific Inc.) e Bioanalyzer 2100 (Agilent Technologies), respectivamente.

3.3. Construção das bibliotecas de cDNA e sequenciamento via RNA-Seq na plataforma Illumina

Para a construção das bibliotecas de cDNA foi utilizado 1,5 µg de RNA total extraído das folhas para o isolamento e purificação do RNAm, utilizando-se o kit de preparação de amostras TruSeq RNA Sample Preparation v2 (Illumina). As bibliotecas de cDNA foram construídas de acordo com a recomendação do fabricante e o sequenciamento dos cDNAs pelo método de RNA-Seq foi realizado nas plataformas HiScanSQ System e HiSeq 2500 (Illumina), seguindo as instruções do fabricante para sequenciamento “paired-end” de fragmentos de 50 pb e de 100 pb respectivamente.

3.4. Processamento dos dados para identificação de SNPs candidatos

Todas as análises de bioinformática realizadas neste estudo estão resumidas no fluxograma a seguir (Figura 6).

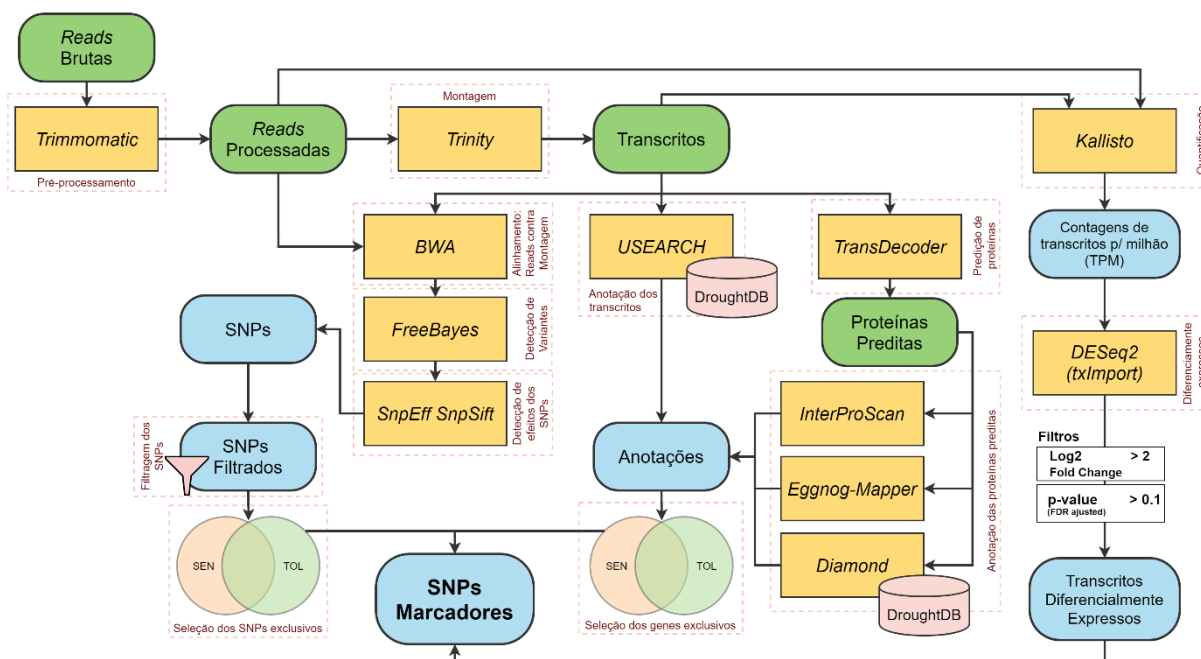


Figura 6. Resumo da metodologia utilizada para as análises de bioinformática após sequenciamento via RNA-Seq.

3.4.1. Pré-processamento das leituras e montagem *de novo* dos transcritos

A partir do sequenciamento dos cDNAs foram obtidas as leituras brutas “reads” no formato FASTQ, as quais passaram por um pré-processamento através do programa Trimmomatic (v.0.36.) (Bolger et al., 2014), para remoção de adaptadores e contaminantes, sequências menores que um tamanho pré-determinado, sequências de baixa qualidade e nucleotídeos não reconhecidos (‘Ns’). Os parâmetros utilizados nesse programa foram: janela de qualidade - 4:5; poda abrupta da ponta 5’ (“leading”): 5 bases; poda abrupta da ponta 3’ (“trailing”): 5 e tamanho mínimo de sequências (“minlen”) de 25. As leituras obtidas a partir da amostra SM6_rep3 levantaram suspeita de contaminação durante teste de agrupamento e por isso não foram utilizadas na montagem.

Apenas as bibliotecas filtradas, sem adaptadores e com regiões de alta qualidade foram utilizadas para realizar a montagem *de novo* do transcriptoma com o programa Trinity (v.2.6.6) (Grabherr et al., 2011). Uma única montagem reuniu o transcriptoma das duas cultivares utilizadas neste estudo (SP81-3250/tolerante e RB85-5453/sensível). Para isso, os seguintes parâmetros foram alterados no Trinity: cobertura mínima de k-mers para que o contig seja montado pela etapa *Inchworm*: -

min_kmer_cov alterado para 3; *-min_glue*, que determina o número mínimo de leituras necessárias para união de duas ou mais sequências contíguas foi aumentado para 3, e o tamanho mínimo de sequências contíguas foi de 300. Os demais parâmetros de configuração do Trinity foram mantidos para a realização da montagem *de novo*.

3.4.2. Análise de expressão diferencial por cultivar, nível e tempo de estresse

Os transcritos obtidos na montagem *de novo* foram quantificados a partir do alinhamento das leituras contra a montagem utilizando o programa Kallisto (v.0.44.0) (Bray et al., 2016), que gerou contagens brutas na unidade de transcritos por milhão (TPM), obtendo-se desse modo uma estimativa de abundância de cada transcrito.

Essas contagens foram utilizadas como entrada no programa DESeq2 (v.1.14.1) (Love et al., 2014), com o pacote txlImport (Soneson et al., 2015) para obter o cálculo dos transcritos diferencialmente expressos (DEs). Foram realizadas todas as comparações possíveis considerando as três variantes (cultivar, intensidade do déficit hídrico e data de avaliação) em relação ao seu equivalente controle (sem déficit hídrico), obtendo assim o resultado da diferença de expressão para cada par de fatores na unidade de “log2FoldChange” (FC).

Foram considerados diferencialmente expressos apenas aqueles cuja razão de módulo de FC foi maior que 2, e de significância estatística (p-valor) menor que 0,05 (usado para as análises exploratórias) e p-valor ajustado abaixo de 0,1. O ajuste é necessário considerando que foram realizados múltiplos teste de hipótese utilizando o método de Benjamini-Hochberg, com base na taxa de falsas descobertas, ou “False Discovery Rate” (FDR) (Benjamini e Hochberg, 1995).

3.4.3. Anotação dos transcritos

Para a caracterização dos transcritos, foi utilizado o programa UBLAST do USEARCH11 para anotação contra os seguintes genomas obtidos do PGSB (Plant Genome and Systems Biology) (Spannagl et al., 2015): *Aegilops tauschii*, *Arabidopsis thaliana*, *Hordeum vulgare*, *Oryza sativa*, *Sorghum bicolor* e *Zea mays*, além do genoma monoploide mosaico da cana-de-açúcar do CIRAD (Garsmeur et al., 2018).

A anotação contra o recente genoma poliploide da cana-de-açúcar (Souza et al., 2019) foi realizada implementando o algoritmo *BlastX* do alinhador Diamond (Buchfink et al., 2015), utilizando a flag *--more-sensitive* e selecionando apenas o melhor alinhamento para cada transcrito.

Para anotação da montagem, considerada então referência, os nucleotídeos foram alinhados com *BlastX* através do programa Blast+ (Camacho et al., 2009), e em seguida o HMMScan (Finn et al., 2011) foi utilizado para predição de domínios proteicos. A partir desses resultados, foi realizada a predição de proteínas com o programa Transdecoder (Haas et al., 2013), considerando como critérios de seleção para proteínas apenas aquelas com ao menos 900 nucleotídeos, utilizando o parâmetro *-retain_long_orfs 900*, ou com alinhamentos no Blast+ *-retain_blastp_hits* ou HMMER *-retain_pfam_hits*, retornando apenas a melhor proteína por transcrito através do parâmetro *-single_best_orf*.

Para anotação das proteínas utilizamos o pacote de programas do InterProScan (Finn et al., 2017). Também foi realizada a anotação baseada em ortólogos, utilizando o programa Egnog-mapper (Huerta-Cepas et al., 2016), com parâmetros baseados em Xiong (2006), que são: identidade mínima de 50% e o “e-value” mínimo de 0,0001. Por fim, foi realizada uma anotação contra o banco de dados DroughtDB (Alter et al., 2015), utilizado como referência para estresse hídrico. Para isso utilizamos o programa Diamond (Buchfink et al., 2015).

3.4.4. Alinhamento e detecção de bases variantes

Os sítios variantes foram identificados a partir do alinhamento das sequências contra a montagem do transcriptoma. O alinhamento de cada biblioteca pré-processada se deu pelo programa BWA (Burrows-Wheeler Aligner) (Li e Durbin, 2009), utilizando o algoritmo *BWA-mem*, com parâmetros pré-determinados (“default”), e posteriormente, os arquivos de alinhamento foram ordenados e indexados com as ferramentas *-sort* e *-index*, respectivamente, do programa SAMTOOLS (Li et al., 2009).

Em seguida, os arquivos de alinhamento foram submetidos ao programa FreeBayes (Garrison e Marth, 2012), que aceita informações de organismos não-

diploides para a predição de variações (“Variant calling”), tais como SNPs, inserções e deleções (“Indels”) e outras mais complexas. As variações detectadas foram filtradas quanto a qualidade e cobertura, sendo no mínimo aceitas aquelas acima do valor 25, baseado em estudo de Cardoso-Silva et al. (2014).

3.4.5. Predição dos efeitos das variações

Com base no transcriptoma de referência, obteve-se o arquivo de coordenadas gênicas, que possibilitou a inferência de efeitos das mutações em regiões específicas. Para isso, o programa TransDecoder (Haas et al., 2013) juntamente com os programas Blastp (Altschul et al., 1997), com o banco de dados UniRef90 (Suzek et al., 2007) e HMM search com o banco de dados Pfam-A (El-Gebali et al., 2018), foram utilizados para predizer as regiões codificantes e não-codificantes dos transcritos na montagem referência.

Os efeitos das variações preditas foram verificados utilizando os parâmetros “default” do programa SnpEff (Cingolani et al., 2012), que fornece informações a respeito dos possíveis efeitos dos sítios polimórficos identificados e do impacto dessas alterações, além de caracterizá-las de acordo com a região genômica em que se inserem.

Adicionalmente, utilizou-se o programa SnpSift (Cingolani et al., 2012) para avaliar a natureza da troca de bases quanto a quantidade de transições (troca entre pirimidinas ou entre purinas) e transversões (troca de uma pirimidina por purina ou vice-versa) encontradas em cada cultivar.

3.4.6. Seleção de variantes marcadores

Em organismos poliploides, a detecção de SNPs é confundida por sequências homólogas, altamente similares devido aos polimorfismos dos subgenomas (Clevenger et al., 2018), que devem ser diferenciados dos SNPs verdadeiros. Portanto, para a detecção de marcadores, foram aplicados filtros, através de scripts feitos “in house” utilizando o software estatístico R (Team R Core, 2013), baseados nas frequências dos SNP putativos ao longo das bibliotecas, de forma a identificar

somente aquelas variações relacionadas com o cultivar e, assim, removendo possíveis variações causadas por mutações somáticas, oriundas de variabilidade individual ou mesmo potenciais falsos-positivos. Os scripts desenvolvidos para aplicação desses filtros estão disponíveis no seguinte endereço: <https://doi.org/10.5281/zenodo.3784875>.

Os filtros foram aplicados em dois níveis, a fim de considerar apenas: (1) aquelas alterações que ocorriam em mais de metade das réplicas biológicas de cada condição e destas, (2) apenas aquelas que ocorriam em mais de metade das bibliotecas da cultivar relacionada. Assim houve uma redução considerável do conjunto inicial de variações. Este conjunto reduzido, foi explorado quanto ao seu possível papel na tolerância ou suscetibilidade ao estresse hídrico, por meio da relação do impacto da variação com o gene/via funcional no qual ocorreram.

3.4.7. Categorização funcional e análise de enriquecimento com base nos termos de ontologia gênica

Através das anotações com os programas InterProScan e Eggno-mapper (item 3.4.3) obtivemos os termos GO (Gene Ontology), que permitiram relacionar as categorias gênicas funcionais associadas aos transcritos anotados. Para visualizar a distribuição dos transcritos nas diferentes categorias funcionais, utilizamos o programa WEGO (Web Gene Ontology Annotation Plot) (Ye et al., 2018). Essa ferramenta, nos permitiu verificar quais as categorias funcionais que estão sendo mais afetadas pela presença de SNPs, ao comparar estatisticamente os transcritos com SNPs de ambas as cultivares, dentro de cada categoria. Além disso, para cada cultivar, comparamos os transcritos com SNPs contra todo universo de transcritos para avaliar se as proporções se mantinham constantes ou se determinadas categorias eram menos afetadas pela presença de SNPs.

4. RESULTADOS E DISCUSSÃO

4.1. Qualidade do material, sequenciamento das bibliotecas de cDNA via RNA-Seq e processamento das leituras brutas

As amostras de RNA obtidas a partir das folhas +1 das plantas de cana-de-açúcar foram isoladas e avaliadas quanto à pureza e integridade das amostras. Todos os procedimentos foram descritos em estudos já publicados do nosso grupo de pesquisa (Belesini et al., 2017; Telles et al., 2019).

O sequenciamento das bibliotecas na plataforma Illumina HiScanSQ System (2x50 pb) e HiSeq2500 (2x100 pb) gerou um total de 2.367.679.551 leituras brutas, sendo 1.712.145.471 leituras de 50 pb e 655.534.080 leituras de 100 pb. Um total de 770 sequências referentes a amostra SM6_rep3 foram excluídas devido a uma possível contaminação, resultando em 2.367.678.781 leituras.

Essas leituras brutas passaram por um pré-processamento para remoção de sequências curtas e de baixa qualidade, adaptadores e contaminantes, resultando no total de 1.798.266.112 leituras que foram utilizadas para realizar a montagem *de novo*.

4.2. Montagem *de novo*

A montagem *de novo* foi realizada conjuntamente para ambas cultivares utilizando o programa Trinity (v.2.6.6) (Grabherr et al., 2011). O processo de montagem gerou um total de 251.653 transcritos com comprimento médio de 1.089 pb, um N50 igual a 1.565 pb e 49,90% de conteúdo de GC. Cardoso-Silva et al. (2014) obtiveram resultados semelhantes na montagem *de novo* de 6 genótipos de cana-de-açúcar, os quais são utilizados como progenitores em programas de melhoramento no Brasil. De acordo com os autores, o comprimento médio dos transcritos foi de 921 pb, o N50 igual a 1.367 pb e o conteúdo GC de 46,39%.

4.3. Anotação dos transcritos

O conjunto de 251.653 transcritos montados foi alinhado contra os genomas

obtidos do PGSB (Spannagl et al., 2015), no intuito de estabelecer a anotação funcional baseada em sequências homólogas. Além disso, também foi realizado o alinhamento contra os genomas de cana-de-açúcar monoploide (Garsmeur et al., 2018) e poliploide recém-publicado (Souza et al., 2019). O resultado do alinhamento dos 251.653 transcritos montados é apresentado na tabela 2.

Tabela 2. Anotação dos transcritos por referência.

| Referência, programa ou pipeline de anotação | Transcritos anotados |
|--|----------------------|
| Genoma de <i>A. thaliana</i> | 33.084 |
| Genoma de <i>H. vulgare</i> | 71.671 |
| Genoma de <i>O. sativa</i> | 73.424 |
| Genoma de <i>A. tauschii</i> | 77.512 |
| Genoma de <i>Z. mays</i> | 90.613 |
| Genoma de <i>S. bicolor</i> | 101.876 |
| Genoma monoploide de <i>Saccharum</i> | 101.128 |
| Genoma poliploide de <i>Saccharum</i> | 92.096 |
| DroughtDB (proteínas) | 4.722 |
| InterProScan | 104.534 |
| Diamond (BlastX) NR | 228.272 |

Os alinhamentos contra os genomas de outras gramíneas como *Z. mays* e *S. bicolor* foram os que mais retornaram resultados de anotação, provavelmente pelo fato desses organismos estarem mais próximos filogeneticamente do gênero *Saccharum*. Entretanto, o número de alinhamentos gerados contra o genoma monoploide e poliploide de cana-de-açúcar recém-publicados (Garsmeur et al., 2018; Souza et al., 2019) foi menor quando comparado ao alinhamento contra o genoma de *S. bicolor*, mesmo pertencendo ao mesmo gênero. Isso pode demonstrar que apesar dos esforços para a obtenção de um genoma completo para a cana-de-açúcar, a alta complexidade e ploidia do seu genoma dificultam a sua exploração.

4.4. Identificação dos SNPs putativos

Após o alinhamento das reads contra a montagem, utilizada então como referência, a identificação dos polimorfismos foi realizada individualmente para cada

cultivar, tolerante e sensível, através do *software* FreeBayes, onde foram obtidos um total de 222.147 variações. A frequência com que as variações ocorrem nos transcritos é de uma variação a cada 1.518 pb para a cultivar tolerante enquanto que, para a cultivar sensível, esta frequência é de uma variação a cada 1.346 pb. De acordo com Cardoso-Silva et al. (2014), as variações exclusivas de cada genótipo apresentam maior probabilidade de estarem envolvidas com características agronômicas de interesse.

Foi observada uma taxa de transição por transversão (TS/TV) de 1,879 para a cultivar sensível e 1,886 para a cultivar tolerante. Essa razão é geralmente pouco variável entre plantas de uma mesma espécie. Assim, essa informação torna-se útil como indicador de qualidade de dados, além de possibilitar o estabelecimento de estudos filogenéticos entre sequências. Em cana-de-açúcar não há tal descrição disponível na literatura, o que provavelmente se deve à dificuldade de estudos genômicos com esse poliploide. De qualquer maneira, é usual que transições superem transversões na maioria dos casos, devido ao fato das primeiras serem mais toleradas pela seleção natural, pois a tendência para gerar mutações não sinônimas em sequências codificadoras está mais relacionada ao número de transversões (Wakeley, 1996; Keller et al., 2007).

Para manter apenas os polimorfismos relacionados aos genótipos, foram aplicados filtros (descritos em material e métodos - item 3.4.6) a fim de eliminar variações biológicas individuais. Com isso, obtivemos um total de 26.012 polimorfismos putativos, divididos entre exclusivos da cultivar tolerante, exclusivos da cultivar sensível e comum para ambas cultivares, conforme ilustrado na Figura 7. Todos os resultados obtidos antes e após filtragem dos dados estão disponíveis no seguinte endereço: <https://doi.org/10.5281/zenodo.3784875>.

O *software* SnpEff foi utilizado para anotar os 26.012 polimorfismos filtrados, que foram divididos em cinco categorias: (1) polimorfismos de nucleotídeo único (SNP); (2) polimorfismos de múltiplos nucleotídeos (MNP) - que ocorrem quando há troca de mais de um nucleotídeo em sequência; (3) inserções; (4) deleções e (5) variações complexas - quando ocorrem dois ou mais polimorfismos diferentes simultaneamente.

As cinco categorias foram contabilizadas separadamente para as duas

cultivares e os resultados encontram-se na Figura 8. Os resultados foram semelhantes para ambas, com predominância de variações do tipo SNP. Esse resultado já era esperado uma vez que as plantas são evolutiva e filogeneticamente próximas. Além disso, os polimorfismos do tipo SNP são os mais comuns dentro do genoma das plantas (Rafalski, 2002).

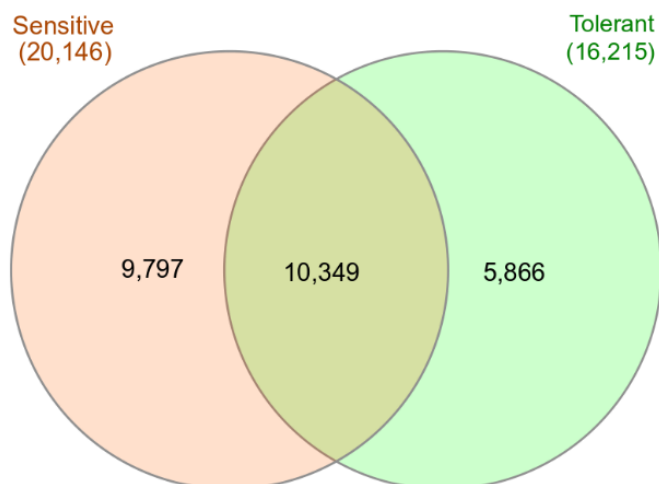


Figura 7. Número total de variações encontradas exclusivamente em cada cultivar (Sensível e Tolerante) e em comum para ambas, considerando apenas aquelas que ocorrem em mais de metade das réplicas biológicas de cada condição e destas, apenas aquelas que ocorrem em mais de metade das bibliotecas da cultivar relacionada.

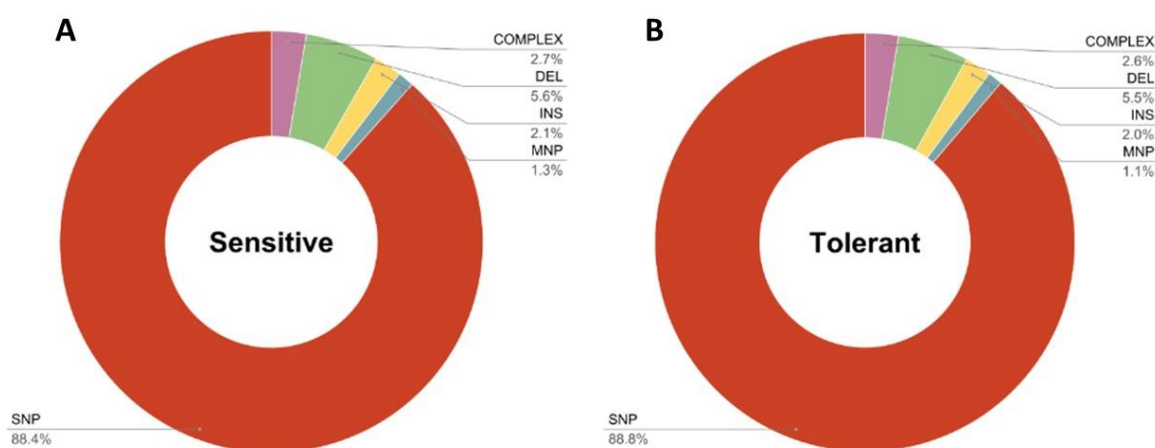


Figura 8. Distribuição das diferentes categorias de polimorfismos nas duas cultivares de cana-de-açúcar: (A) aqueles que ocorrem exclusivamente na cultivar sensível; (B) aqueles que ocorrem exclusivamente na cultivar tolerante. SNP - polimorfismo de nucleotídeo único; MNP - polimorfismo de múltiplos nucleotídeos; INS - inserção; DEL - deleção e COMPLEX - variações complexas.

Ainda com o software SnpEff, foi feita a predição da distribuição desses polimorfismos nas regiões gênicas para cada cultivar (Figura 9). Aqui o resultado também foi semelhante para as duas cultivares, com aproximadamente $\frac{2}{3}$ das variações encontradas em regiões não codificadoras e $\frac{1}{3}$ em regiões codificadoras. As mutações que ocorrem em regiões não codificadoras do genoma são muito mais numerosas, como é de se esperar, pelo fato de a maior fração do genoma ser representada por DNA não codificador. Essas mutações são consideradas mutações neutras pois, em quase sua totalidade, não afetam produtos gênicos ou a expressão dos genes (Alberts et al., 2017).

Entretanto, não podemos ignorar completamente a importância dessas regiões não codificantes, visto que diversos estudos mostraram que esses trechos podem apresentar diversas outras funções além da síntese proteica (Eddy, 2001; Liu et al., 2012; Romero-Barrios et al., 2018).

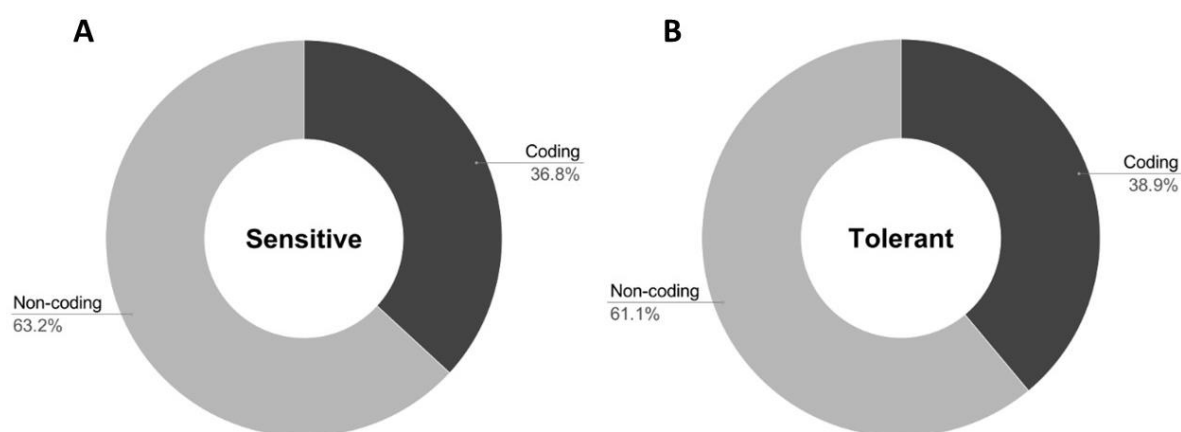


Figura 9. Distribuição dos polimorfismos de acordo com a posição que ocorrem no gene. (A) cultivar sensível; (B) cultivar tolerante.

As variações também foram classificadas em relação ao impacto que poderiam causar na proteína (Tabela 3). O tipo de impacto mais predominante é o modificador, causado por variações que ocorrem geralmente em regiões não codificadoras. Pelo mesmo motivo exposto anteriormente, sugerimos que essas mutações possam ter sofrido menor pressão de seleção durante o melhoramento da cana-de-açúcar devido à menor magnitude de seus efeitos, por isso ocorrem em maior proporção.

As variações de impacto alto representaram menos de 2% do total de

polimorfismos detectados, provavelmente pelo motivo oposto às variações de impacto moderado. Como aponta Ahloowalia et al. (2004), é pouco provável que a mudança radical da estrutura de uma proteína acarrete em uma maior eficiência metabólica, uma vez que esse fato depende de um ajuste global em uma rota bioquímica, na qual estão envolvidas diversas moléculas e seus genes precursores.

Entretanto, ao longo do processo de melhoramento da cana-de-açúcar para obtenção de genótipos com características favoráveis, a seleção fenotípica pode ter resultado no acúmulo gradativo de mutações favoráveis e, conseqüentemente, houve a eliminação de mutações com efeito deletério, seja pela seleção natural ou artificial.

Menegatto (2017) realizou um estudo de caracterização de polimorfismos em cana-de-açúcar através da técnica de genotipagem-por-sequenciamento, utilizando dados genotípicos de indivíduos do Painel Brasileiro de Genótipos de Cana-de-Açúcar como população. Com auxílio do software SnpEff, o autor realizou a predição das possíveis funções e posições dos polimorfismos. Foi observado que variações de impacto alto apareciam com menor frequência. Entretanto, após fixação dos alelos polimórficos, ou seja, considerando apenas aqueles alelos com polimorfismos que apareciam em no mínimo 1% da população, os genótipos melhorados apresentaram maior proporção de mutações com impacto alto.

Com isso, pode ser possível que as variações encontradas em nosso trabalho sejam evolutivamente favoráveis, pois, ao longo do melhoramento da cana-de-açúcar, essas variações foram positivamente selecionadas no genoma como um todo. Contudo, estudos em populações são necessários para tal comprovação.

Tabela 3. Classificação das variações em dados brutos e filtrados nas cultivares sensível e tolerante à estresse hídrico de acordo com a magnitude predita de seu efeito. Os sítios variantes foram classificados pelo SnpEff nas categorias de impacto alto, moderado e modificador.

| Impacto previsto | Sensível | Tolerante |
|-------------------------|-----------------|------------------|
| Alto | 375 (1,86%) | 307 (1,89%) |
| Moderado | 7.049 (34,99%) | 6.008 (37,04%) |
| Modificador | 12.724 (63,15%) | 9.904 (61,06%) |

As variações com impacto alto na proteína são, em maioria, mutações de mudança de quadro (“Frameshift variant”) e/ou mutações de sentido trocado

(“Missense variant”) (Figura 10). A primeira consiste na inserção ou exclusão de um número de pares de bases que não é um múltiplo de três, interrompendo o quadro de leitura tripla da sequência de DNA. Geralmente levam à criação de um códon de terminação prematuro (códon de parada) e resultam em um produto proteico truncado (mais curto que o normal). Já as mutações de sentido trocado são causadas pela substituição de um único par de bases, alterando o código genético de uma maneira que produz um aminoácido diferente do aminoácido usual nessa posição, podendo alterar ou não a função da proteína (Cingolani et al., 2012); Robert e Pelletier, 2018).

Com relação às variações de impacto moderado, encontramos predominantemente mutações de sentido trocado (“Missense variant”) e/ou inserção de códons entre e dentro de outros códons (“conservative inframe insertion” e “disruptive inframe insertion”, respectivamente) (Figura 11). Mesmo sendo encontrados com maior frequência, a justificativa é similar ao que foi exposto para as mutações de impacto alto.

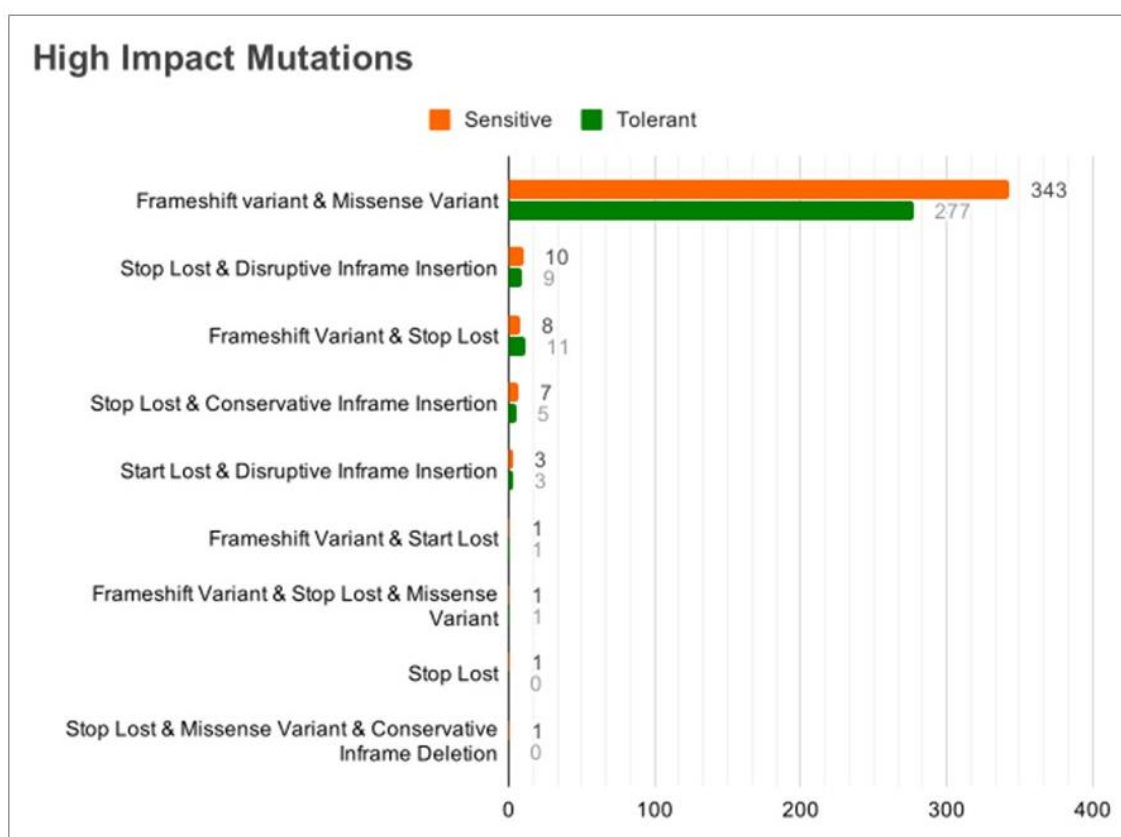


Figura 10. Efeitos preditos dos polimorfismos nos genes, classificados em impacto alto segundo classificação de Cingolani et al. (2012).

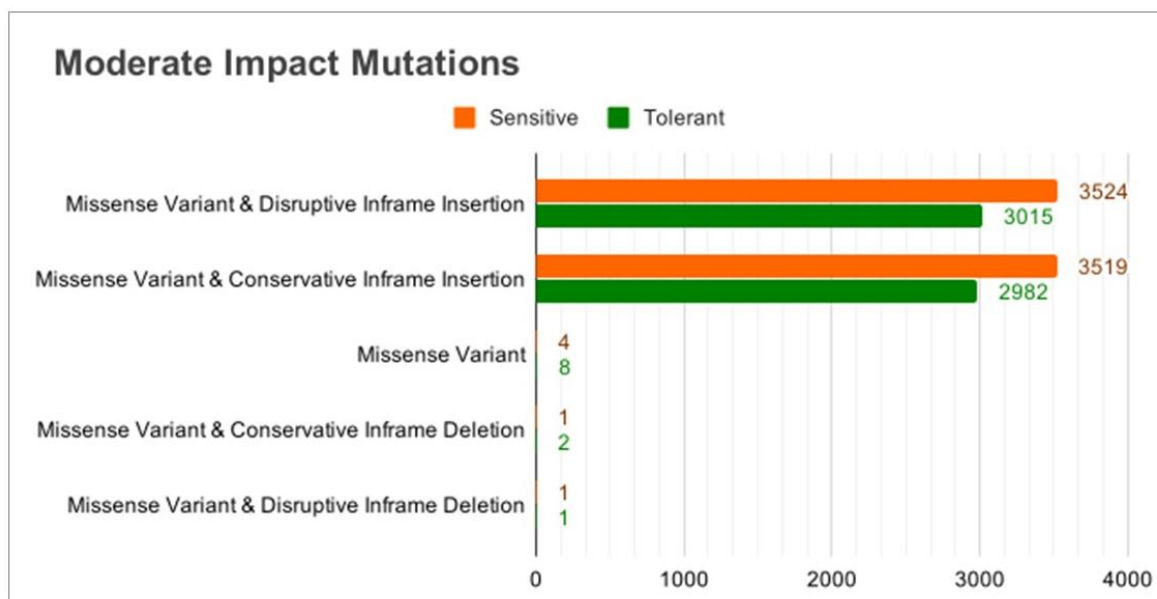


Figura 11. Efeitos preditos dos polimorfismos nos genes, classificados em impacto moderado segundo classificação de Cingolani et al. (2012).

4.5. Categorização funcional e análise de enriquecimento com base nos termos de ontologia gênica

Para analisar as possíveis respostas funcionais das duas cultivares de cana-de-açúcar aos tratamentos de estresse hídrico, os genes que apresentaram SNPs putativos foram analisados sobre a perspectiva de processos biológicos possibilitada pelo enriquecimento de termos de ontologia gênica (“Gene Ontology”). Essa análise foi realizada com o programa WEGO, que permite criar um gráfico através de uma classificação fornecida pelo “Gene Ontology”, facilitando assim a visualização da anotação (Ye et al., 2018). O programa utiliza o teste exato de Fisher (p -valor $<0,05$), com uma correção de falsas descobertas (FDR) em termos de processos biológicos e funções moleculares.

A anotação de ontologias gênicas baseia-se em mineração de dados oriundos de sequências públicas com anotação disponível, com similaridade mínima de 60% entre as sequências. O resultado da análise é classificado em três domínios: a) processo biológico, que indica as operações ou os conjuntos de eventos moleculares com início e fim definidos em que o produto do gene estudado participa e que são fundamentais para o funcionamento de uma célula, tecido, órgãos ou organismo; b)

função molecular, que indica as atividades elementares do produto do gene a nível molecular; e c) componente celular, que indica a parte da célula ou o ambiente extracelular em que o produto do gene desempenha sua função (Ashburner et al., 2000).

Um total de 1.990 genes receberam anotação com base nos termos de ontologia gênica. As cultivares foram comparadas entre si afim de obtermos as proporções de cada uma com relação ao número de SNPs. Na grande maioria das categorias, a cultivar sensível apresentou maior proporção de genes com SNPs em relação a cultivar tolerante, com exceção de alguns termos relacionados à fotossíntese, como membrana fotossintética, membrana do tilacoide e atividade da enzima hidrolase (Figura 12).

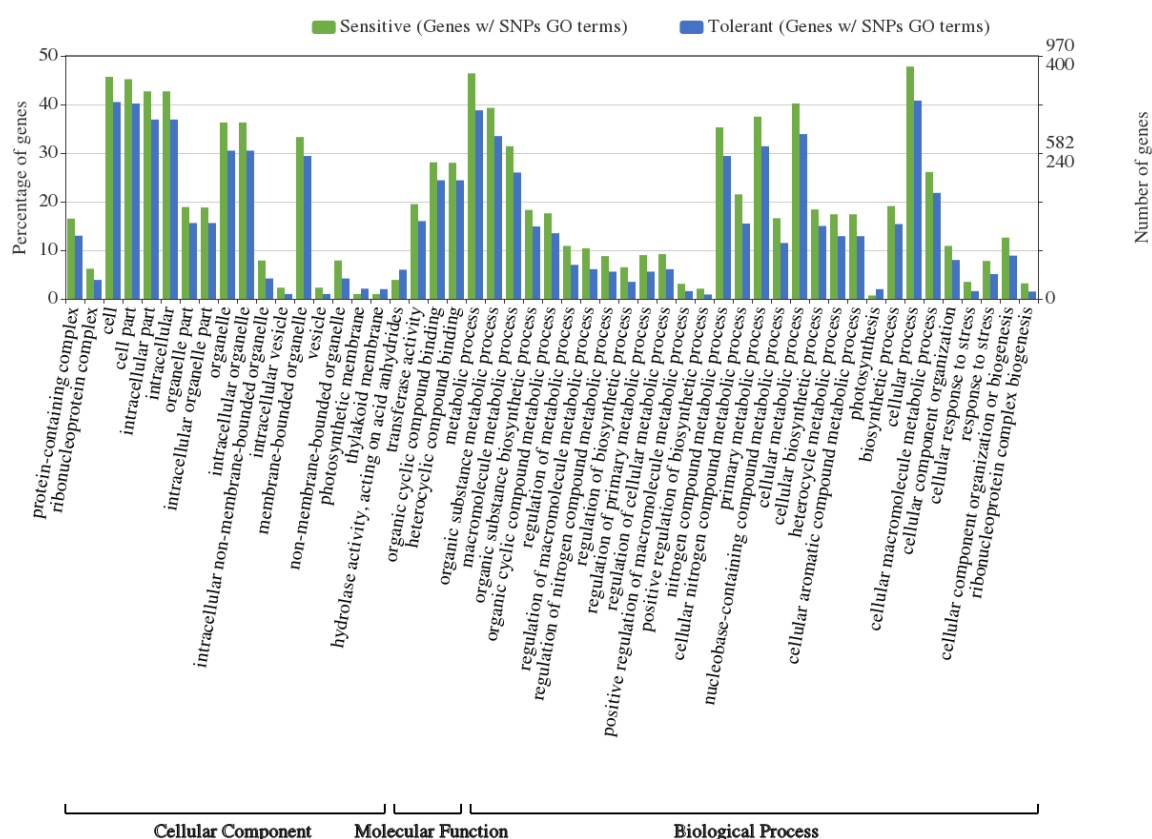


Figura 12. Categorização funcional e análise de enriquecimento com base nos termos de ontologia gênica realizada através do programa WEGO. As barras verdes representam genes com SNPs exclusivos da cultivar sensível e barras azuis são aqueles genes com SNPs exclusivos da cultivar tolerante.

Em trabalho realizado anteriormente pelo nosso grupo de pesquisa, Telles (2016) avaliou a taxa de fotossíntese das mesmas cultivares do nosso trabalho -

SP81-3250 (tolerante) e RB85-5453 (sensível) sob estresse hídrico moderado, severo e controle em cada época amostral de 30, 60 e 90 dias. A autora observou que a atividade fotossintética na cultivar tolerante foi maior que na sensível aos 90 dias de estresse, na condição de déficit hídrico severo. Com isso, sugerimos que a diferença entre as proporções de termos anotados relacionados à fotossíntese possa estar associada a maior eficiência da cultivar tolerante frente aos efeitos do estresse hídrico.

4.6. Genes associados a SNPs putativos relacionados à resposta ao estresse hídrico

O estresse hídrico, assim como os demais estresses abióticos, atua como um sinal primário que é percebido pelas células, desencadeando uma cascata de sinalização que leva à ativação de fatores de transcrição e induz a transcrição de genes específicos, cujos produtos funcionais estão relacionados com o restabelecimento da homeostasia celular, proteção e reparo das proteínas e membranas danificadas e eliminação de radicais livres e compostos tóxicos (Shinozaki e Yamaguchi-Shinozaki, 2000; Hoekstra et al., 2001).

Aqui destacamos algumas variações que apresentaram impacto alto ou moderado em genes que codificam proteínas relacionadas com a resposta ao estresse hídrico nas plantas. Essas variações ocorreram exclusivamente na cultivar tolerante e apenas em regiões codificantes dos genes. Os SNPs putativos presentes nessas regiões podem alterar a estrutura primária (sequência), a conformação (estrutura secundária e terciária) e conseqüentemente, a função da proteína. Dessa forma, esperamos que essas variações estejam relacionadas a um aumento no potencial da planta em reagir a situações de estresse, conferindo caráter de tolerância à planta.

Também avaliamos a expressão gênica dos genes que apresentaram SNPs, e todos os genes aqui destacados tiveram sua expressão alterada em decorrência do estresse aplicado. Além disso, alguns desses genes apresentaram expressão diferencial quando comparadas ao controle, o que reforça o indicativo de sugerirmos esses genes como potenciais candidatos a marcadores moleculares para seleção assistida. Todos esses resultados estão resumidos na tabela 4.

Tabela 4. SNPs candidatos presentes em genes de resposta ao estresse hídrico. **ID:** identificação do gene; **POS:** posição do SNP na sequência; **REF:** nucleotídeo de referência; **ALT:** nucleotídeo alterado pelo SNP.

| ID | POS | REF | ALT | Aminoácido trocado | Efeito predito | Impacto | Anotação PGSB | Anotação genoma novo | Expressão diferencial |
|-------------|------|-----|-----|--------------------|---|----------|-------------------------|-------------------------|---------------------------|
| 40843_c4_g1 | 817 | A | G | Ser → Ser | Mutação de sentido trocado e inserção de códon conservativa | Moderado | CDPK | NA | NA |
| 34949_c1_g1 | 1540 | C | T | Gli → Gli | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | LRR-RPK | NA | NA |
| | 2462 | C | T | Ser → Asn | Mutação de sentido trocado e inserção de códon conservativa | Moderado | LRR-RPK | NA | NA |
| 36275_c1_g1 | 1299 | C | G | Ser → Arg | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Serina/Treonina quinase | Serina/Treonina quinase | Severo 90d x Controle 90d |
| | 1500 | G | A | Lis → Lis | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Serina/Treonina quinase | Serina/Treonina quinase | Severo 90d x Controle 90d |
| | 1539 | G | A | Arg → Arg | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Serina/Treonina quinase | Serina/Treonina quinase | Severo 90d x Controle 90d |
| 48440_c6_g2 | 1385 | G | A | Asp → Asp | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |
| 46100_c3_g1 | 1250 | C | T | Ala → Ala | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |
| 46278_c3_g1 | 1355 | C | T | Ser → Ser | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |
| 42279_c2_g1 | 1340 | G | T | Pro → Pro | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |
| | 1426 | A | G | Leu → Leu | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Serina/Treonina quinase | NA | NA |
| | 1431 | G | A | Pro → Leu | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Serina/Treonina quinase | NA | NA |
| | 1436 | T | A | Arg → Arg | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |
| | 1607 | T | C | Ser → Ser | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |
| | 1691 | C | G | Glu → Asp | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Serina/Treonina quinase | NA | NA |

Continuação da tabela 4.

| ID | POS | REF | ALT | Aminoácido trocado | Efeito predito | Impacto | Anotação PGSB | Anotação genoma novo | Expressão diferencial |
|-------------|------|--------|--------|--------------------|---|----------|--|----------------------------------|-----------------------------|
| 41837_c1_g2 | 286 | A | G | Lis → Glu | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de ligação do elemento responsivo ao ABA (ABF) | Elemento responsivo ao ABA | Moderado 90d x Controle 90d |
| | 1008 | G | C | Ala → Gli | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de resposta a auxina (ARF) | NA | Severo 30d x Controle 30d |
| 37325_c0_g1 | 889 | C | T | Val → Met | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de resposta a auxina (ARF) | NA | Severo 30d x Controle 30d |
| | 924 | G | A | Ala → Val | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de resposta a auxina (ARF) | NA | Severo 30d x Controle 30d |
| | 2144 | T | C | Leu → Pro | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de resposta a auxina (ARF) | Fator de resposta a auxina (ARF) | NA |
| 43317_c2_g1 | 2308 | A | G | Ile → Val | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de resposta a auxina (ARF) | Fator de resposta a auxina (ARF) | NA |
| | 2480 | A | T | Gln → Leu | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de resposta a auxina (ARF) | Fator de resposta a auxina (ARF) | NA |
| 48302_c0_g1 | 4561 | C | A | Val → Leu | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de resposta a auxina (ARF) | NA | NA |
| 32452_c0_g1 | 462 | TGT GC | TCT GC | His → Asp | Mudança no quadro de leitura e mutação de sentido trocado | Alto | Fator de transcrição MYB | NA | NA |
| 47341_c0_g1 | 87 | G | A | Arg → His | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de transcrição MYB | MYB | NA |
| | 1690 | T | A | Ile → Ile | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de transcrição MYB | MYB | Severo 30d x Controle 30d |
| | 1756 | A | G | Pro → Pro | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de transcrição MYB | MYB | Severo 30d x Controle 30d |
| 36247_c2_g1 | 1888 | G | A | Ala → Ala | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de transcrição MYB | MYB | Severo 30d x Controle 30d |
| | 1921 | A | G | Pro → Pro | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de transcrição MYB | MYB | Severo 30d x Controle 30d |
| | 2072 | G | T | Ala → Ser | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de transcrição MYB | MYB | Severo 30d x Controle 30d |
| | 2144 | A | C | Arg → Arg | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de transcrição MYB | MYB | Severo 30d x Controle 30d |

Continuação da tabela 4.

| ID | POS | REF | ALT | Aminoácido trocado | Efeito predito | Impacto | Anotação PGSB | Anotação genoma novo | Expressão diferencial |
|-------------|------|----------|----------|--------------------|---|----------|---|--|-----------------------------|
| 39011_c2_g1 | 502 | C | G | Pro → Ala | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de transcrição MYB | NA | NA |
| 48115_c0_g2 | 670 | A | G | Val → Ala | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de transcrição MYB | MYB | NA |
| 45438_c1_g3 | 192 | G | A | Ala → Thr | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fatores de Transcrição Dedos de Zinco (Zinc Finger) | Dedos de Zinco (Zinc Finger) | NA |
| 45824_c2_g1 | 1029 | CG CC | CG CT | Ala → Ala | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Proteína de Choque Térmico (HSP) | Proteína de Choque Térmico (HSP) | NA |
| | 1140 | C | T | Gli → Gli | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Proteína de Choque Térmico (HSP) | Proteína de Choque Térmico (HSP) | NA |
| | 606 | C | T | Asn → Asn | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Proteína de Choque Térmico (HSP) | Proteína de Choque Térmico (HSP) | NA |
| 43402_c0_g1 | 2641 | A | G | Asn → Ser | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de transcrição de estresse térmico (HSF) | Fator de transcrição de estresse térmico (HSF) | Severo 90d x Controle 90d |
| | 3106 | C | T | Ala → Val | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Fator de transcrição de estresse térmico (HSF) | Fator de transcrição de estresse térmico (HSF) | Severo 90d x Controle 90d |
| 33867_c2_g1 | 659 | C | G | Cis → Ser | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Fator de transcrição de estresse térmico (HSF) | NA | NA |
| 45604_c3_g2 | 1504 | C | T | Tre → Met | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | Glutaciona S-transferase (GST) | NA | NA |
| | 689 | T | G | Tre → Glu | Perda do códon de parada e inserção de códon disruptiva | Alto | NA | Glutaciona S-transferase (GST) | NA |
| 35415_c1_g1 | 364 | C | A | Tre → Tre | Mutação de sentido trocado e inserção de códon conservativa | Moderado | NA | Glutaciona S-transferase (GST) | NA |
| | 442 | G | T | Gln → His | Mutação de sentido trocado e inserção de códon conservativa | Moderado | NA | Glutaciona S-transferase (GST) | NA |
| 29783_c0_g1 | 339 | A | G | Ile → Val | Mutação de sentido trocado e inserção de códon disruptiva | Moderado | NA | Glutaciona S-transferase (GST) | NA |
| 46450_c0_g2 | 350 | G | A | Ala → Val | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Proteína Universal de Estresse (USP) | NA | Moderado 90d x Controle 90d |
| | 360 | G | T | Pro → Tre | Mutação de sentido trocado e inserção de códon conservativa | Moderado | Proteína Universal de Estresse (USP) | NA | Moderado 90d x Controle 90d |

4.6.1. Proteínas quinases dependentes de cálcio (CDPKs), Serina/Treonina quinases (Ser/Tre) e Receptores do tipo quinase com repetições ricas em leucina (LRR-RPK)

As proteínas quinases estão envolvidas na regulação de aspectos-chaves de diversas funções celulares, incluindo divisão celular, metabolismo e respostas a sinais externos. As proteínas quinases catalisam a transferência de γ -fosfato de ATP (ou GTP) para os substratos proteicos, processo conhecido como fosforilação. A fosforilação/desfosforilação de proteínas tem sido reconhecida como um dos principais mecanismos da regulação pós-traducional da atividade de proteínas e da transdução de sinais intracelulares em organismos eucarióticos. Dessa forma, a fosforilação pode resultar na ativação ou inativação de proteínas (Cheng et al., 2011).

O Ca^{2+} presente no citoplasma funciona como um mensageiro secundário, responsável pela ativação de diversas proteínas quinases que regulam a expressão gênica (Taiz et al., 2017). As proteínas quinases dependentes de Ca^{2+} (CDPKs) implementam os sinais de Ca^{2+} no citoplasma por meio da fosforilação, desempenhando ação em cascatas de sinalização, induzidas, por exemplo, por frio, seca, salinidade, ferimentos e deficiência nutricional (White e Broadley, 2003). Diversos estudos comprovam a participação das CDPKs na tolerância ao estresse hídrico em plantas (Ma e Wu, 2007; Xu et al., 2010; Wei et al., 2014).

Semelhante a muitas proteínas quinases, as quinases receptoras (RPKs) são proteínas transmembranares e estão envolvidas em múltiplos processos fisiológicos, incluindo desenvolvimento, resistência a doenças, percepção hormonal e respostas ao estresse (Ouyang et al. 2010). Quando a planta é submetida a condições de estresse abiótico, as RPKs percebem os ligantes extracelulares e ativam a via a jusante através da fosforilação dos domínios intracelulares de serina/treonina quinase (Osakabe et al. 2013).

As RPKs podem ser divididas em três subclasses principais de acordo com a especificidade dos diferentes substratos entre os domínios quinases. As subclasses são os (1) receptores tirosina quinase, que estão envolvidos na fosforilação de resíduos de tirosina; (2) receptores serina treonina quinase, que fosforilam resíduos de serina e treonina; e (3) receptores histidina quinase que fosforilam resíduos de

histidina.

Segundo Afzal et al. (2008) os receptores serina treonina quinases interagem com outras proteínas para realizar uma ampla gama de processos, desempenhando um papel central na sinalização durante o reconhecimento do estresse, a subsequente ativação dos mecanismos de defesa e o controle do desenvolvimento das plantas. De acordo com Hong et al. (1997), o LRR-RPK é induzido por ABA, desidratação, alto teor de sal e baixas temperaturas.

Neste trabalho, encontramos 15 variações de impacto moderado do tipo SNP em 7 genes diferentes que codificam quinases. O gene 36275_c1_g1 que codifica uma proteína serina/treonina quinase apresentou três SNPs em posições diferentes, sendo um deles do tipo transversão, provocado pela troca de uma citosina por uma guanina, resultando na troca dos aminoácidos de Serina para Arginina (Tabela 4). Esse gene também apresentou expressão diferencial induzida sob estresse hídrico severo aos 90 dias de tratamento quando comparado ao controle do mesmo período, o que indica que ele pode estar atuando em conjunto com uma cascata de sinalização em resposta ao caráter de tolerância à seca. Além disso, esse gene recebeu anotação quando alinhado ao novo genoma da cana-de-açúcar recém-publicado (Souza et al., 2019).

4.6.2. Fatores de Transcrição ABF, ARF, MYB e Zinc Finger

Uma das mais importantes respostas às condições ambientais adversas é o controle da transcrição. Os fatores de transcrição (FTs) funcionam na ativação ou repressão gênica através da associação de domínios de ligação a sequências consenso do DNA na região do promotor de genes alvo (Akhtar et al., 2012).

Muitos genes que codificam fatores de transcrição são induzidos por estresse, sugerindo que vários mecanismos de regulação da transcrição podem funcionar no ajuste das vias de transdução dos sinais de estresse provocado por seca, frio ou alta salinidade (Shinozaki e Yamaguchi-Shinozaki, 2007). Geralmente esses FTs são agrupados em famílias com base no seu domínio de ligação ao DNA, incluindo CBF/DREB, ABF/ABAE, HSF, bZIP, MYC/MYB e Zinc finger (Zhu, 2001; 2002).

Os FTs aqui citados apresentam fundamental importância nas plantas por

estarem envolvidos em respostas aos estresses abióticos e abióticos, especialmente a seca. Esses FTs possivelmente estão atuando na ativação de muitos genes relacionados aos mecanismos de resposta ao déficit hídrico em cana-de-açúcar. Assim, sugerimos que os polimorfismos presentes nestes genes estejam contribuindo de alguma forma para o caráter de tolerância dessa planta.

4.6.2.1. Fator de ligação do elemento responsivo ao ABA (ABF)

As plantas sintetizam compostos do metabolismo secundário, denominados fitormônios, para adequar seu metabolismo às diversas condições ambientais e regular os processos fisiológicos necessários ao seu desenvolvimento, sendo o ácido abscísico (ABA) um dos mais importantes. É bem estabelecido na literatura que a seca desencadeia a produção de ABA, o qual resulta no fechamento dos estômatos e na indução da expressão de genes relacionados ao estresse (Yamaguchi-Shinozaki e Shinozaki, 2005).

Alguns genes regulados por ABA compartilham um elemento *cis* comum de 8 a 10 pb conhecido como ABA-responsive element (ABRE). Um pequeno grupo de fatores de transcrição bZIP (*basic leucine zipper*), AREB (*ABA responsive element binding protein*) e ABF (*ABA responsive element binding factor*) podem se ligar a ABRE e ativar a expressão dos genes relacionados à tolerância ao estresse hídrico (Busk e Pagès, 1998).

Plantas mutantes de algodão superexpressando os genes *GhABF2D* e *AtABF3* mostraram maior resistência em condições de déficit hídrico reduzindo a perda de água por transpiração através do fechamento dos estômatos (Kerr et al., 2018). Neste trabalho, um gene anotado como fator de ligação do elemento responsivo ao ABA apresentou uma variação de impacto moderado do tipo SNP de transição, provocando a troca de uma adenina por uma guanina, com subsequente alteração do aminoácido Lisina para Glutamato, o que pode resultar em uma proteína diferente da esperada (Tabela 4).

Além disso, esse gene também apresentou expressão diferencial sob estresse moderado aos 90 dias de tratamento quando comparado com o controle no mesmo período (Tabela 4). Tanto a variação quanto a expressão diferencial foram observadas

exclusivamente na cultivar tolerante, indicando um possível candidato a marcador molecular para resistência à seca.

4.6.2.2. Fator de resposta a auxina (ARF)

Outro fitormônio relacionado com a resposta das plantas ao estresse hídrico é a auxina, também denominado ácido indolacético (AIA). O AIA foi um dos primeiros fitormônios identificados em plantas, onde é predominantemente sintetizado em tecidos com crescimento rápido, como meristemas apicais, folhas jovens, frutos e sementes em desenvolvimento (Ljung et al., 2001). Aloni et al. (2006) também relataram que as extremidades das raízes podem sintetizar esse composto, regulando a sua distribuição. Além disso, o AIA desempenha um papel chave na regulação do crescimento das plantas, como o alongamento das células, desenvolvimento de tecidos vasculares e dominância apical (Wang et al., 2001).

Embora o AIA tenha sido amplamente reconhecido por seu papel na regulação do crescimento e desenvolvimento da planta, esse fitormônio pode ainda regular a resposta do crescimento da planta sob estresses abióticos, inclusive seca (Jain e Khurana, 2009; Eyidogan et al., 2012). Ribaut e Pilet (1994) utilizaram raízes de milho submetidas ao déficit hídrico para avaliar se mudanças nos níveis de auxina durante o estresse. Os autores observaram que os níveis de auxina subiam com o aumento no nível de estresse hídrico e concluíram que a auxina pode apresentar um importante papel na regulação do metabolismo e crescimento das raízes em condições de estresse hídrico.

A expressão de AIA é controlada, basicamente, por duas famílias de fatores de transcrição: ARF e AUX/AIA (Guilfoyle e Hagen, 2007). Os ARFs (fatores de resposta a auxina) são capazes de se ligar a elementos de resposta TGTCTC presentes nos promotores de genes que respondem a auxina, ativando ou reprimindo suas transcrições ou ainda recrutando outros fatores de transcrição (Guilfoyle e Hagen, 2007), que mediam a resposta a auxina (Song et al., 2009). Com isso, foi observado que os ARFs estão relacionados à sinalização hormonal (Li et al., 2006), senescência (Ellis et al., 2005) e desenvolvimento (Krogan et al., 2012).

Kang et al. (2018) avaliaram os efeitos do gene *IbARF5* de batata doce em

plantas transgênicas de *Arabidopsis* e observaram que houve um aumento na tolerância à seca e salinidade. Essas plantas apresentaram aumento no conteúdo de carotenoides, teor de ABA, prolina e aumento na atividade da enzima superóxido dismutase, com conseqüente diminuição no conteúdo de H_2O_2 . Os carotenoides são um dos precursores do ABA e apresentam um papel funcional no desenvolvimento e adaptação ambiental das plantas (Schwartz et al., 2003). Dessa forma, sugere-se que um aumento no conteúdo de carotenoides proporciona uma melhor adaptação das plantas a condições adversas.

Neste trabalho foram encontrados três genes que codificam ARFs com sete variações de impacto moderado do tipo SNP, sendo três transversões e quatro transições. Todos os sete SNPs provocaram troca de aminoácidos (Tabela 4), o que pode alterar a proteína final. O gene 37325_c0_g1 apresentou expressão diferencial sob estresse hídrico severo aos 30 dias de tratamento quando comparado ao controle do mesmo período, o que pode estar relacionado com uma resposta precoce da planta em condições de estresse hídrico severo. Além disso, o gene 43317_c2_g1_i2 também recebeu anotação no alinhamento contra o novo genoma da cana-de-açúcar recém-publicado (Souza et al., 2019).

4.6.2.3. Fatores de transcrição MYB

Os fatores de transcrição MYB estão presentes em todos os eucariotos e são uma das maiores famílias de FTs conhecidas em plantas. Eles estão associados à regulação de diversos processos nas plantas, como metabolismo secundário, respostas a hormônios e regulação de morfogênese celular, além da participação na defesa contra fatores bióticos e abióticos (Buck e Atchley, 2003; Dubos et al., 2010).

A síntese de MYB nas plantas acompanha o acúmulo de ABA endógeno durante os estágios mais avançados de respostas ao estresse. Esta via também pode ser ativada por ácido jasmônico (JA) que é produzido pela planta em resposta a estresses bióticos e ferimentos (Abe et al., 2003). Segundo Lindemose et al. (2013), a interação de genes da família de FTs MYB com o ABA resulta diversos mecanismos de resposta ao estresse abiótico, tais como regulação da abertura estomática, supressão da ativação do gene responsivo ao ácido jasmônico, regulação da resposta

ao estresse por seca através da interação entre ABA e auxina, entre outros.

Outros trabalhos comprovam a participação dos FTs MYB na resposta das plantas ao estresse hídrico em diferentes espécies de plantas, inclusive cana-de-açúcar (Cominelli et al., 2005; Rodrigues et al., 2011; Shan et al., 2012; Chen et al., 2015). Com relação aos polimorfismos, Parida et al. (2012) validaram 362 SNPs relacionados com estresse biótico e abiótico em 91 genótipos de arroz. Desse total de variações, 10 SNPs ocorreram em genes que codificam FTs da família MYB.

Neste trabalho foram encontrados cinco genes que codificam fatores de transcrição da família MYB e que apresentaram variações de impacto alto e moderado do tipo SNP, sendo que um deles (36247_c2_g1) apresentou 6 variações diferentes no mesmo gene (Tabela 4). Além disso, esse gene apresentou expressão diferencial sob estresse hídrico severo aos 30 dias de tratamento quando comparado ao controle do mesmo período. Os SNPs e a expressão diferencial presente nesse gene podem, provavelmente, estar relacionadas à resposta ou tolerância ao estresse hídrico na cana-de-açúcar.

O SNP do gene 32452_c0_g1 provocou uma variação de impacto alto na proteína entre os genes discutidos. Essa variação foi classificada como SNP de transversão, com a troca do nucleotídeo Guanina para Citosina, causando uma mudança no quadro de leitura e substituição do aminoácido Histidina para Asparagina.

4.6.2.4. Fatores de Transcrição Dedos de Zinco (“Zinc Finger”)

As proteínas “Zinc Finger” estão entre as mais abundantes nos genomas eucarióticos, apresentando funções diversas que incluem reconhecimento de DNA, embalagem de RNA, regulação da apoptose, dobragem e montagem de proteínas, ligação lipídica e ativação transcricional (Laity et al., 2001). Essas proteínas apresentam alta afinidade para ligação com sequências específicas de DNA e possuem domínios de ligação ao DNA característicos em fatores de transcrição (Margolin et al., 1994).

Os fatores de transcrição “Zinc Finger” (ZF) compõem uma família de fatores de transcrição consideravelmente grande em plantas, podendo ser classificadas em diferentes tipos de proteínas, como C2H2, C2HC, C2HC5, C3HC4, CCCH, C4,

C4HC3, C6 e C8, baseado no número e localização dos resíduos de cisteínas e histidinas ligadas a um íon de zinco na estrutura secundária do dedo (“finger”) (Seok et al., 2016).

Os ZF podem desempenhar importantes funções biológicas pela interação com o DNA ou cromatina, RNA e outras proteínas, além de estarem envolvidos com o crescimento e desenvolvimento de plantas, respostas a fitormônios e a estresses bióticos e abióticos (Li et al., 2013). Em plantas, o tipo C2H2 está envolvido no crescimento e desenvolvimento das mesmas. De acordo com Kielbowicz-Matuk (2012), existe um potencial na manipulação de genes C2H2 para conferir tolerância a salinidade, frio, seca, calor, estresse osmótico e oxidativo.

Neste trabalho, um gene anotado como ZF apresentou uma variação de impacto moderado do tipo SNP, onde ocorreu uma transição com a troca de uma guanina por uma adenina, resultando na substituição do aminoácido Alanina por Treonina e que pode influenciar na composição da proteína final (Tabela 4). Diversos trabalhos já demonstraram a relação dos zinc finger com mecanismos de tolerância ao estresse (Wang et al., 2008; Peng et al., 2012; Seok et al., 2016), o que pode tornar esse polimorfismo um potencial candidato à marcador molecular para tolerância à seca.

4.6.2.5. Fatores de transcrição de estresse térmico (HSF) e Proteínas de Choque Térmico (HSP)

Os fatores de transcrição de estresse térmico (heat stress transcription factor - HSF) estão envolvidos na regulação da expressão de diversos genes responsivos ao estresse, desempenhando um papel chave na tolerância à vários estresses abióticos, como salinidade, frio e seca. Além disso, alguns estudos comprovam que fitormônios relacionados ao estresse abiótico, como ácido jasmônico (JA), ácido abscísico (ABA), ácido salicílico (SA) e etileno (Et) regulam a expressão de genes HSF (Hu et al., 2015; Huang et al., 2015; Zhang et al., 2015).

Os HSFs se encontram como monômeros na ausência de estresses, porém ao serem submetidos a altas temperaturas, são convertidos para a forma trimerizada e se ligam à região de controle no gene, conhecida como Elemento de Resposta ao

Choque Térmico (HSE) que, por sua vez, induz a expressão das Proteínas de Choque Térmico (HSP) (Guo et al., 2008).

A principal função das HSPs é atuar como chaperonas, evitando a agregação ou desnaturação protéica, além de promover a renaturação das proteínas que perderam sua conformação ativa (Liberek et al., 2008; Tyedmers et al., 2010). Existe uma forte interação entre o dobramento de proteínas celulares e sistemas de degradação mediados por chaperonas. Tem sido observado que HSPs induzidas por déficit hídrico previnem a agregação de proteínas e a desnaturação durante o mesmo (Vaseva et al., 2012).

Neste trabalho, dois genes anotados como HSFs apresentaram três variações de impacto moderado do tipo SNP, provocando duas transições e uma transversão que resultaram na troca dos aminoácidos (Tabela 4). Essas trocas provavelmente resultarão em uma proteína diferente da esperada. Também observamos que o gene 43402_c0_g1 apresentou expressão diferencial sob estresse severo aos 90 dias de tratamento quando comparado com o controle no mesmo período, que vai de acordo com o descrito por Dossa et al. (2016) em estudo do perfil de expressão de genes HSF em plantas de gergelim em resposta ao estresse hídrico, onde 90% dos genes HSF responderam à seca. Com isso, sugerimos que esse gene apresenta potencial para um possível candidato à marcador molecular para resistência à seca.

Também identificamos um gene que codifica uma HSP com três variações de impacto moderado do tipo SNP, todos de transição, porém em nenhum deles houve troca de aminoácidos (Tabela 4). Os genes 45824_c2_g1 e 43402_c0_g1 receberam anotação no alinhamento com o novo genoma da cana-de açúcar recém-publicado.

4.6.3. Glutathione S-transferase

Durante o déficit hídrico pode ocorrer um desequilíbrio entre a eliminação enzimática e não enzimática de Espécies Reativas de Oxigênio (EROs), resultando em um súbito aumento do potencial redox celular, o que pode ocasionar danos a diversos componentes celulares como lipídios, proteínas e ácidos nucleicos (Labudda e Azam, 2014). Assim, a capacidade da planta de inativar as EROs é resultado do perfeito funcionamento de seu metabolismo antioxidante e ajuste osmótico, conferindo

característica de tolerância à seca (Angelova et al., 2000).

O efeito negativo das EROs é chamado estresse oxidativo, podendo causar danos fotoinibitórios e fotooxidativos em plantas sob condições de seca. O sistema antioxidante das plantas é responsável pela eliminação das EROs através de enzimas como superóxido dismutase (SOD), catalase (CAT), ascorbato-peroxidase (APX), glutathione redutase (GR), glutathione peroxidase (GPX), guaiacol peroxidase (GPOX) e glutathione S-transferase (GST) (Asada, 1999).

As glutathione S-transferases são uma família de enzimas de detoxificação bem caracterizadas e que estão envolvidas na tolerância das plantas à estresses abióticos. Essas enzimas catalisam a conjugação da glutathione reduzida (GSH) a substratos eletrofílicos produzindo conjugados destes compostos que são solúveis em água, o que reduz sua toxicidade e protege os tecidos vegetais dos danos causados pelo estresse oxidativo (Anderson e Davis, 2004).

Chen et al. (2012) avaliaram plantas mutantes de *Arabidopsis* superexpressando o gene AtGSTU17 e observaram que houve maior tolerância à seca e ao estresse salino quando comparadas às plantas selvagens. As plantas mutantes também apresentaram maiores níveis de ABA, menores aberturas estomáticas e menores taxas de transpiração.

Gallé et al. (2009) também verificaram a presença de GST em plantas de trigo submetidas ao déficit hídrico durante o período de enchimento dos grãos. Os autores observaram que os genes TaGSTU1B e TaGSTF6 que codificam GSTs apresentaram um alto nível de expressão induzida pela seca em quatro cultivares avaliadas, sendo que a maior expressão foi detectada em uma cultivar tolerante.

Neste trabalho, encontramos um gene que codifica uma GST com uma variação de impacto moderado do tipo SNP de transição, em que ocorreu a troca de uma citosina por uma timina, resultando na substituição do aminoácido Treonina por Metionina. Entretanto, um caso peculiar chamou atenção. Dois genes que receberam anotação apenas contra o novo genoma da cana-de-açúcar recém-publicado (Souza et al., 2019) codificam uma GST. Em um deles, o gene 35415_c1_g1 apresentou uma variação de impacto alto do tipo SNP de transversão, que provocou a perda do códon de parada desta proteína (Tabela 4).

4.6.4. Universal Stress Protein

As USPs foram originalmente descobertas em *Escherichia coli* e posteriormente identificada em diversos outros organismos, como arqueias, bactérias, plantas e metazoários. Como seu próprio nome sugere, essas proteínas participam de uma ampla gama de respostas celulares a estresses bióticos e abióticos (Nyström e Neidhardt, 1992).

As funções das USPs são bastante distintas, podendo estar associadas à eliminação de íons, respostas à hipóxia, mobilidade celular e regulação do crescimento e desenvolvimento celular, além de apresentarem atividade de ligação, reparo e remodelamento do DNA e transporte de proteínas celulares (Kvint et al., 2003; Drumm et al., 2009).

Essas proteínas são superexpressas quando submetidas a condições ambientais desfavoráveis, incluindo falta de nutrientes (deficiência de carbono, nitrogênio, fosfato, sulfato e aminoácidos), estresse por calor/frio, estresse oxidativo, estresse hídrico, toxicidade por metais pesados, exposição a etanol, antibióticos, entre outros (VanBogelen et al., 1990; Kvint et al., 2003).

De acordo com Jung et al. (2015), as USPs apresentam fundamental importância na tolerância a estresses abióticos. A superexpressão do gene USP induz a ativação de alguns genes pouco expressos durante o estresse, o que promove uma melhora no crescimento da planta, diminui o acúmulo de EROs, mantém a homeostase iônica e melhora o estado fisiológico da planta, levando a proteção da célula contra danos oxidativo.

Em estudo realizado por Loukehaich et al. (2012), os autores avaliaram a superexpressão do gene SpUSP (*Solanum pennellii*) em tomate transgênico sob condições de seca, e observaram que houve um aumento nos níveis de ABA e indução do fechamento estomático com consequente redução na perda de água, conferindo característica de tolerância a seca e melhor eficiência fotossintética a planta.

Em outro estudo semelhante, Udawat et al. (2016), avaliaram os efeitos do gene SbUSP (*Salicornia brachiata*) em plantas de tabaco transgênicas submetidas a diferentes condições de estresse abiótico. Os autores concluíram que houve aumento

no teor de clorofila, conteúdo relativo de água, prolina, açúcar total, açúcar redutor, aminoácidos livres, potencial osmótico, conteúdo de polifenóis, estabilidade da membrana, além de menor perda de eletrólitos, menor peroxidação lipídica e menor acúmulo de radicais H_2O_2 .

Neste trabalho, identificamos um gene que codifica uma USP e que apresentou duas variações de impacto moderado do tipo SNP, sendo um de transição e outro de transversão, ambos provocando alteração de aminoácidos e que possivelmente terão impacto na proteína final (Tabela 4).

5. CONCLUSÕES

Os resultados deste estudo indicam que os “softwares” FreeBayes e SnpEff foram eficientes na busca por SNPs candidatos e seus possíveis efeitos em duas cultivares de cana-de-açúcar contrastantes quanto à tolerância ao déficit hídrico. Assim, foi possível identificar SNPs candidatos exclusivos da cultivar tolerante em alguns genes relacionados com a resposta da planta ao estresse hídrico, como fatores de transcrição ABF, ARF, MYB e “Zinc Finger”, proteínas quinases, proteínas de choque térmico, proteína universal do estresse e enzimas de detoxificação como a Glutathione S-transferase, sugerindo uma possível relação desses polimorfismos com a tolerância da planta.

Além disso, os resultados apresentados para a busca e análise de SNPs candidatos para os genes acima descritos mostram que existe uma quantidade limitada de modificações nas regiões codificadoras desses genes. Isso pode ser explicado pela existência de uma pressão de seleção muito forte sobre esses genes ao longo da evolução relacionada com a importância das proteínas correspondentes na cascata de eventos na qual elas estão envolvidas.

A descoberta de novos SNPs para a cultura da cana-de-açúcar é de fundamental importância para os programas de melhoramento, uma vez que podem diminuir o longo tempo necessário para o lançamento de novos cultivares, através do mapeamento de traços de interesse agrônomico por meio da seleção assistida por marcador (MAS). A seleção assistida por marcadores incorpora informações de marcadores moleculares na seleção de genótipos superiores e mostra algumas vantagens quando comparada à seleção fenotípica, que é a mais utilizada nos programas de melhoramento. Dentre elas cabe salientar a seleção de caracteres onde a avaliação fenotípica exige alto investimento, como a resistência ao déficit hídrico por exemplo. Com isso, podemos concluir que os polimorfismos encontrados são potenciais candidatos a marcadores moleculares e que, futuramente, poderão ser utilizados em programas de melhoramento através de seleção assistida por marcadores.

Além disso, a geração desse tipo de conhecimento poderá, futuramente, permitir a aplicação de técnicas de edição gênica visando a obtenção de plantas mais

adaptadas à seca. Por fim, a disponibilização de uma referência gênica em cana-de-açúcar, contribuirá para as pesquisas relacionadas a esta cultura de grande importância econômica.

6. REFERÊNCIAS

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. **The Plant Cell** 15:63-78.

Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. **Molecular Plant-Microbe Interactions** 21:507-517.

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varieties. **Euphytica** 135:187-204.

Ahmad P, Latef A, Arafat AH, Rasool S, Akram, NA, Ashraf M, Guzel S (2016) Role of proteomics in crop stress tolerance. **Frontiers in plant science** 7:1336-1349.

Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, Singh NK (2012) DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. **Journal of Genetics** 91:385–395.

Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, Walter P, Wilson J, Hunt T (2017) *Biologia molecular da célula*. Porto Alegre: Artmed, 1464p.

Aloni R, Aloni E, Langhans M, Ullrich CI (2006) Role of cytokinin and auxin in shaping root architecture: regulating vascular differentiation, lateral root initiation, root apical dominance and root gravitropism. **Annals of botany** 97:883-893.

Alter S, Bader KC, Spannagl M, Wang Y, Bauer E, Schön CC, Mayer KFX (2015) DroughtDB: an expert-curated compilation of plant drought stress genes and their homologs in nine species. **Database** 2015:1-7.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. **Nucleic Acids Research** 25:3389-3402.

Andersen JR, Lübberstedt T (2003) Functional markers in plants. **Trends in plant science** 8:554-560.

Anderson JV, Davis DG (2004) Abiotic stress alters transcript profiles and activity of glutathione S-transferase, glutathione peroxidase, and glutathione reductase in *Euphorbia esula*. **Physiologia plantarum** 120: 421-433.

Angelova MB, Pashova SB, Slokoska LS (2000) Comparison of antioxidant enzyme biosynthesis by free and immobilized *Aspergillus niger* cells. **Enzyme and Microbial Technology** 26:544-549.

Asada K (1999) The water-water cycle in chloroplasts: scavenging of reactive oxygens and dissipation of excess photons. **Annual Review of Plant Physiology and Plant Molecular Biology** 50:601-639.

Ashburner M et al. (2000) Gene Ontology: tool for the unification of biology. **Nature Genetics** 25:25–29.

Assad ED, Pinto HS, Nassar A, Harfuch L, Freitas S, Farinelli B, Lundell M, Bachion LC, Fernandes ECM (2013) **Impacto das mudanças climáticas na produção agrícola brasileira**. Banco Mundial: PROFOR, 116 p.

Atteya AM (2003) Alteration of water relations and yield of corn genotypes in response to drought stress. **Bulgarian Journal of Plant Physiology** 29:63-76.

Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. **Critical reviews in plant sciences** 24:23-58.

Belesini AA, Carvalho FMS, Telles BR, de Castro GM, Giachetto PF, Vantini JS, Carlin SD, Cazetta JO, Oinheiro DG, Ferro MIT (2017) *De novo* transcriptome assembly of sugarcane leaves submitted to prolonged water-deficit stress. **Genetics and Molecular Research** 16:1-20.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal statistical society** 57:289-300.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. **Bioinformatics** 30:2114–2120.

Borém A, Caixeta ET (Eds.) (2009) *Marcadores Moleculares*. Viçosa: Editora UFV, 385p.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American journal of human genetics** 32:314-331.

Bray EA (1993) Molecular responses to water deficit. **Plant Physiology** 103:1035-1040.

Bray NL, Pimentel H, Melsted P e Pachter L (2016) Near-optimal probabilistic rna-seq quantification. **Nature Biotechnology** 34:525-529.

Brookes AJ (1999) The essence of SNPs. **Gene** 234:177-186.

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. **Nature methods** 12:1-5.

Buck MJ, Atchley WR (2003) Phylogenetic analysis of plant basic helix-loop-helix proteins. **Journal of molecular evolution** 56:742-750.

Busk PK, Pagès M (1998) Regulation of abscisic acid-induced transcription. **Plant molecular biology** 37:425-435.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. **BMC bioinformatics** 10:421-429.

Cambráia J (2005) Aspectos bioquímicos, celulares e fisiológicos dos estresses nutricionais em plantas. In: Nogueira RJMC, Araújo EL, Willadino LG, Cavalcante UMT (Eds.) **Estresses Ambientais: Danos e Benefícios em Planta**. Pernambuco: Imprensa Universitária da UFRPE, 127-137.

Cardoso-Silva CB et al. (2014) De novo assembly and transcriptome analysis of contrasting sugarcane varieties. **PLoS One** 9:1-10.

Castro SS, Abdala K, Silva AA, Bôrges VMS (2010) A expansão da cana-de-açúcar no cerrado e no estado de goiás: elementos para uma análise espacial do processo. **Boletim Goiano de Geografia** 30:171-191.

Chaves MM (1991) Effects of water deficits on carbon assimilation. **Journal of experimental Botany** 42:1-16.

Chaves MM, Flexas J, Pinheiro C (2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. **Annals of botany** 103:551-560.

Chaves MM, Maroco JP, Pereira JS (2003) Understanding plant responses to drought—from genes to the whole plant. **Functional plant biology** 30:239-264.

Chen JH, Jiang HW, Hsieh EJ, Chen HY, Chien CT, Hsieh HL, Lin TP (2012) Drought and salt stress tolerance of an *Arabidopsis* glutathione S-transferase U17 knockout mutant are attributed to the combined effect of glutathione and abscisic acid. **Plant Physiology** 158:340-351.

Chen T, Li W, Hu X, Guo J, Liu A, Zhang B (2015) A cotton MYB transcription factor, GbMYB5, is positively involved in plant adaptive response to drought stress. **Plant and Cell Physiology** 56:917-929.

Cheng HC, Qi RZ, Paudel H, Zhu HJ (2011) Regulation and function of protein kinases and phosphatases. **Enzyme Research** 2011:1-3.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. **Fly** 6:80-92.

Clevenger JP, Korani W, Ozias-Akins P, Jackson S (2018) Haplotype-Based Genotyping in Polyploids. **Frontiers in Plant Science** 9:564-569.

Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL, Tonelli C (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. **Current biology** 15:1196-1200.

Conab - Companhia Nacional de Abastecimento (2019): **Acompanhamento da Safra Brasileira - Cana-de-açúcar, safra 2018/2019, quarto levantamento**. Brasília: CONAB. 75 p.

Cordeiro GM, Elliott F, McIntyre CL, Casu RE, Henry RJ (2006) Characterisation of single nucleotide polymorphisms in sugarcane ESTs. **Theoretical and Applied Genetics** 113:331-343.

Costa AM (2011) Prospecção gênica e bioinformática. In: Faleiro FG, Andrade SEM, dos Reis Junior FB (Eds.) **Biotecnologia: estado da arte e aplicações na agropecuária**. Planaltina: Embrapa Cerrados, p. 119-141.

Costa EA et al. (2016) QTL mapping including codominant SNP markers with ploidy level information in a sugarcane progeny. **Euphytica** 211:1-16.

D'Hont A, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. **Molecular and General Genetics MGG** 250:405-413.

D'Hont A, Souza GM, Menossi M, Vincentz M, Van-Sluys MA, Glaszmann JC, Ulian E (2008) Sugarcane: a major source of sweetness, alcohol, and bio-energy. In: Moore PH, Ming R (Eds.) **Genomics of tropical crop plants**. New York: Springer, p. 483-513.

Danecek P et al. (2011) The variant call format and VCFtools. **Bioinformatics** 27:2156-2158.

Daniels J, Roach BT (1987) Taxonomy and evolution. In: Heinz DJ (Ed.) **Sugarcane improvement through breeding**. Amsterdam: Elsevier, p. 7-84.

Devarumath RM, Kalwade SB, Kulkarni PA, Sheelavanthmath SS, Suprasanna P (2013) Integrating omics approaches in sugarcane improvement. In: Barh D (Ed.) **OMICS Applications in Crop Science**. Boca Raton: CRC Press, p. 191-250.

Dossa K, Diouf D, Cissé N (2016) Genome-wide investigation of Hsf genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. **Frontiers in plant science** 7:1522-1538.

Drumm JE et al. (2009). *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. **PLoS Pathogens** 5:1-13.

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. **Trends in plant science** 15:573-581.

Eddy SR (2001) Non-coding RNA genes and the modern RNA world. **Nature Reviews Genetics** 2:919-929.

Edwards GE, Huber SC (1981) The C4 pathway. In: Hatch MD, Boardman NK (Eds.) **Photosynthesis. The biochemistry of plants, a comprehensive treatise**. New York: Academic Press, p. 237-281.

El-Gebali S (2018) The Pfam protein families database in 2019. **Nucleic acids research** 47:427-432.

Ellis CM, Nagpal P, Young JC, Hagen G, Guilfoyle TJ, Reed JW (2005) Auxin response factor1 and auxin response factor2 regulate senescence and floral organ abscission in *Arabidopsis thaliana*. **Development** 132:4563-4574.

Endres L, Silva JV, Ferreira VM, Barbosa GVS (2010) Photosynthesis and water relations in Brazilian sugarcane. **The Open Agriculture Journal** 4:31-37.

Eyidogan F, Oz MT, Yucel M, Oktem HA (2012) Signal transduction of phytohormones under abiotic stresses. In: Khan N, Nazar R, Iqbal N, Anjum N (Eds.) **Phytohormones and Abiotic Stress Tolerance in Plants**. Berlin: Springer, p. 1-48.

Faleiro FG, Andrade SEM (2011) Biotecnologia: uma visão geral. In: Faleiro FG, Andrade SEM, dos Reis Junior FB (Eds.) **Biotecnologia: estado da arte e aplicações na agropecuária**. Planaltina: Embrapa Cerrados, p. 13-30.

Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. Brasília: EMBRAPA-CENARGEN, 220p.

Ferreira TH, Gentile A, Vilela RD, Costa GGL, Dias LI, Endres L, Menossi M (2012) microRNAs associated with drought response in the bioenergy crop sugarcane (*Saccharum* spp.). **PLoS One** 7:1-14.

Ferro MIT (2008) **Estudo do estresse hídrico em cana-de-açúcar utilizando genômica funcional**. 212 f. Tese (Livre Docência) - Unesp, Jaboticabal.

FIESP (2018) **Outlook Fiesp 2028: projeções para o agronegócio brasileiro**. São Paulo: FIESP, 86 p.

Finn RD et al. (2017) InterPro in 2017-beyond protein family and domain annotations. **Nucleic acids research** 45:190-199.

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. **Nucleic acids research** 39:29-37.

Fragoso RR, Nakasu EYT, Rocha TL, Rosa AJM (2011) Genômica funcional. In: Faleiro FG, Andrade SEM, dos Reis Junior FB (Eds.) **Biotecnologia: estado da arte e aplicações na agropecuária**. Planaltina: Embrapa Cerrados, p. 143-173.

Freire LP (2010) **Estudo da expressão e análise de polimorfismos de genes candidatos para a tolerância à seca em cafeeiro**. 118 f. Dissertação (Mestrado em Biotecnologia Vegetal) - UFLA, Lavras.

Fu X et al. (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. **BMC Genomics** 10:161-169.

Gallé A, Csiszár J, Secenji M, Guóth A, Cseuz L, Tari I, Györgyey J, Erdei L (2009) Glutathione transferase activity and expression patterns during grain filling in flag leaves of wheat genotypes differing in drought tolerance: response to water deficit. **Journal of Plant Physiology** 166:1878-1891.

Galmés J, Ribas-Carbó M, Medrano H, Flexas J (2011) Rubisco activity in Mediterranean species is regulated by the chloroplastic CO₂ concentration under water stress. **Journal of Experimental Botany** 62:653-665.

Garcia AA et al. (2013) SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. **Scientific reports** 3:1-10.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. **arXiv** 1:1-9.

Garsmeur O et al. (2018) A mosaic monoploid reference sequence for the highly complex genome of sugarcane. **Nature Communications** 9:1-10.

Gascho GJ, Shih SF (1983) Sugarcane. In: Teare ID, Peet MM (Eds.) **Crop Water Relations**. New York: John Wiley & Sons, p. 445-479.

Gentile A et al. (2013) Effects of drought on the microtranscriptome of field-grown sugarcane plants. **Planta** 237:783-798.

Giachetto PF, Higa RH (2014) Bioinformática aplicada à agricultura. In.: Massruhá SMFS, Leite MAA, Luchiari Junior A, Romani LAS (Eds.) **Tecnologias da informação e comunicação e suas relações com a agricultura**. Brasília: EMBRAPA, p 67-84.

Gibson NJ (2006) The use of real-time PCR methods in DNA sequence variation analysis. **Clinica Chimica Acta** 363:32-47.

Grabherr MG et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature biotechnology** 29:644-652.

Guilfoyle TJ, Hagen G (2007) Auxin response factors. **Current opinion in plant biology** 10:453-460.

Guo L, Chen S, Liu K, Liu Y, Ni L, Zhang K, Zhang L (2008) Isolation of heat shock factor hsf1a-binding sites in vivo revealed variations of heat shock elements in *Arabidopsis thaliana*. **Plant and Cell Physiology** 49:1306-1315.

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science** 80:524-535.

Gupta V et al. (2010) The water-deficit stress-and red-rot-related genes in sugarcane. **Functional & integrative genomics** 10:207-214.

Haas BJ et al. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols** 8:1494–1512.

Hamada H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. **Proceedings of the National Academy of Sciences** 79:6465-6469.

Hara K, Watabe H, Sasazaki S, Mukai F, Mannen H (2010) Development of SNP markers for individual identification and parentage test in a Japanese Black cattle population. **Animal science journal** 81:152-157.

Hatch MD (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. **Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics** 895:81-106.

Henry RJ (2012) *Molecular markers in plants*. Ames: Wiley-Blackwell, 216p.

Hoekstra FA, Golovina EA, Buitink J (2001) Mechanisms of plant desiccation tolerance. **Trends in plant science** 6:431-438.

Hong SW, Jon JH, Kwak JM, Nam HG (1997) Identification of a receptor-like protein kinase gene rapidly induced by abscisic acid, dehydration, high salt, and cold treatments in *Arabidopsis thaliana*. **Plant physiology** 113:1203-1212.

Hsiao TC, Acevedo E (1974) Plant responses to water deficits, efficiency, and drought resistance. **Developments in Agricultural and Managed Forest Ecology** 1:59-84.

Hu Y, Han YT, Wei W, Li YJ, Zhang K, Gao YR, Zhao FL, Feng JY (2015). Identification, isolation, and expression analysis of heat shock transcription factors in the diploid woodland strawberry *Fragaria vesca*. **Frontiers in plant science** 6:736-751.

Huang Y, Li MY, Wang F, Xu ZS, Huang W, Wang GL, Ma J, Xiong AS (2015). Heat shock factors in carrot: genome-wide identification, classification, and expression profiles response to abiotic stress. **Molecular biology reports** 42:893-905.

Huerta-Cepas J et al. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic acids research** 44:286-293.

Jain M (2012) Next generation sequencing technologies for gene expression profiling in plants. **Briefings in functional genomics** 2:63-70.

Jain M, Khurana JP (2009) Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. **The FEBS journal** 276:3148-3162.

Jamaux I, Steinmetz A, Belhassen E (1997) Looking for molecular and physiological markers of osmotic adjustment in sunflower. **New Phytologist** 137:117-127.

Jung YJ, Melencion SMB, Lee ES, Park JH, Alinapon CV, Oh HT, Yun DJ, Chi YH, Lee SY (2015). Universal stress protein exhibits a redox-dependent chaperone function in *Arabidopsis* and enhances plant tolerance to heat shock and oxidative stress. **Frontiers in plant science** 6:1-11.

Kahl G, Mast A, Tooke N, Shen R, van den Boom D (2005) Single nucleotide polymorphisms: detection techniques and their potential for genotyping and genome mapping. In.: Meksem K, Kahl G (Eds.) **The Handbook of Plant Genome Mapping: Genetic and Physical Mapping**. Weinheim: Wiley VCH, p. 75-107.

Kang C, He S, Zhai H, Li R, Zhao N, Liu Q (2018) A sweetpotato auxin response factor gene (IbARF5) is involved in carotenoid biosynthesis and salt and drought tolerance in transgenic *Arabidopsis*. **Frontiers in plant science** 9:1-13.

Keller I, Bensasson D, Nichols RA (2007) Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. **PLoS Genetics** 3:1-7.

Kerr TC, Abdel-Mageed H, Aleman L, Lee J, Payton P, Cryer D, Allen RD (2018) Ectopic expression of two AREB/ABF orthologs increases drought tolerance in cotton (*Gossypium hirsutum*). **Plant, cell & environment** 41:898-907.

Kiełbowicz-Matuk A (2012) Involvement of plant C2H2-type zinc finger transcription factors in stress responses. **Plant Science** 185:78-85.

Konrad D (2019) **Dinâmica do perfil transcricional de duas cultivares de cana-de-açúcar contrastantes à seca e submetidas ao déficit hídrico prolongado**. 94 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Unesp, Jaboticabal.

Krogan NT, Ckurshumova W, Marcos D, Caragea AE, Berleth T (2012) Deletion of MP/ARF5 domains III and IV reveals a requirement for Aux/IAA regulation in *Arabidopsis* leaf vascular patterning. **New Phytologist** 194:391–401.

Kumar LS (1999) DNA markers in plant improvement: an overview. **Biotechnology advances** 17:143-182.

Kvint K, Nachin L, Diez A, Nyström T (2003) The bacterial universal stress protein: function and regulation. **Current opinion in microbiology** 6:140–145.

Labudda M, Azam FMS (2014) Glutathione-dependent responses of plants to drought: a review. **Acta Societatis Botanicorum Poloniae** 83:1-10.

Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. **Current opinion in structural biology** 11:39-46.

Larcher W (2000) **Ecofisiologia vegetal**. São Carlos: Rima, 531p.

Lawlor DW, Cornic G (2002) Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. **Plant, cell & environment** 25:275-294.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics** 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map Format and SAMtools. **Bioinformatics** 25:2078–2079.

Li J, Dai X, Zhao Y (2006) A role for auxin response factor 19 in auxin and ethylene signaling in *Arabidopsis*. **Plant physiology** 140:899-908.

Li WT, He M, Wang J, Wang YP (2013) Zinc finger protein (ZFP) in plants-A review. **Plant Omics** 6:474-480.

Liberek K, Lewandowska A, Ziętkiewicz S (2008) Chaperones in control of protein disaggregation. **The EMBO journal** 27:328-335.

Lindemose S, O'Shea C, Jensen MK, Skriver K (2013) Structure, function and networks of transcription factors involved in abiotic stress responses. **International journal of molecular sciences** 14:5842-5878.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. **The Plant Cell** 24:4333-4345.

Ljung K, Bhalerao RP, Sandberg G (2001) Sites and homeostatic control of auxin biosynthesis in *Arabidopsis* during vegetative growth. **The plant journal** 28:465-474.

Loukehaich R, Wang T, Ouyang B, Ziaf K, Li H, Zhang J, Lu Y, Ye, Z (2012) SpUSP, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. **Journal of experimental botany** 63:5593-5606.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology** 15:1-21.

Ma SY, Wu WH (2007) AtCPK23 functions in Arabidopsis responses to drought and salt stresses. **Plant molecular biology** 65:511-518.

Machado RS, Ribeiro RV, Marchiori PER, Machado DFSP, Machado EC, Landell MGA (2009) Respostas biométricas e fisiológicas ao déficit hídrico em cana-de-açúcar em diferentes fases fenológicas. **Pesquisa agropecuária brasileira** 44:1575-1582.

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. **Archives of biochemistry and biophysics** 444:139-158.

Marconi TG, Costa EA, Miranda HRCAN, Mancini MC, Cardoso-Silva CB, Oliveira KM, Pinto LR, Mollinari M, Garcia AAF, Souza AP (2011) Functional markers for gene mapping and genetic diversity studies in sugarcane. **BMC Research Notes** 4:1-9.

Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, Rauscher FJ (1994) Krüppel-associated boxes are potent transcriptional repression domains. **Proceedings of the National Academy of Sciences**. 91:4509–4513.

Marguerat S, Bähler J (2010) RNA-seq: from technology to biology. **Cellular and molecular life sciences** 67:569-579.

Margulies M et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. **Nature** 437:376-381.

Marin FR. Variedades de cana-de-açúcar. **Ageitec** - Agência Embrapa de Informação Tecnológica. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_42_1110200717570.html> . Acesso em: 15 set 2019.

Markad NR, Kale AA, Pawar BD, Jadhav AS, Patil SC (2014) Molecular characterization of sugarcane (*Saccharum officinarum* L.) genotypes in relation to salt tolerance. **The Bioscan** 9:1785-1788.

Matsuoka S, Ferro J, Arruda P (2009) The Brazilian experience of sugarcane ethanol industry. **In Vitro Cellular & Developmental Biology-Plant** 45:372-381.

Matsuoka S, Garcia AAF, Arizono H (2005) Melhoramento da cana-de-açúcar. In.: Borém A (Ed.) **Melhoramento de espécies cultivadas**. Viçosa: UFV, p. 205-251.

McCord P, Glynn N, Comstock J (2019) Identifying markers for resistance to sugarcane orange rust (*Puccinia kuehni*) via selective genotyping and capture sequencing. **Euphytica** 215:1-14.

McKendry P (2002) Energy production from biomass (part 1): overview of biomass. **Bioresource technology** 83:37-46.

Menegatto LS (2017) **Caracterização de polimorfismos e assinaturas de seleção entre genótipos de cana-de-açúcar (*Saccharum spp.*) através de genotipagem-por-sequenciamento**. 96 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - USP, Piracicaba.

Moreira LM, Varani AM (2015) Plasticidade e fluxo genômico. In.: Moreira LM (Organizador) **Ciências genômicas: fundamentos e aplicações**. Ribeirão Preto: Sociedade Brasileira de Genética, p. 101-116.

Nelson DL, Cox MM (2014) Princípios de bioquímica de Lehninger. Porto Alegre: Artmed, 1250p.

Nepomuceno AL, Neumaier N, Farias JRB, Oya T (2001) Tolerância à seca em plantas. **Biotecnologia Ciência e Desenvolvimento** 23:12-18.

Nogales A, DeDiego ML (2019) Host single nucleotide polymorphisms modulating influenza a virus disease in humans. **Pathogens** 8:168-189.

Nyström T, Neidhardt FC (1992) Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. **Molecular microbiology** 6:3187-3198.

Osakabe Y, Osakabe K, Shinozaki K, Tran LSP (2014) Response of plants to water stress. **Frontiers in plant science** 5:1-8.

Osakabe Y, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP (2013) Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. **Journal of experimental botany** 64:445-458.

Ouyang SQ, Liu YF, Liu P, Lei G, He SJ, Ma B, Zhang WK, Zhang JS, Chen SY (2010) Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (*Oryza sativa*) plants. **The Plant Journal** 62:316-329.

Parida SK, Mukerji M, Singh AK, Singh NK, Mohapatra T (2012). SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. **BMC genomics** 13:1-17.

Peng X, Zhao Y, Cao J, Zhang W, Jiang H, Li X, Ma Q, Zhu S, Cheng B (2012) CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments. **PLoS one** 7:1-15.

Pimentel C (2004) A relação da planta com a água. Rio de Janeiro: Seropédica, 191p.

Pincelli RP (2010) **Tolerância à deficiência hídrica em cultivares de cana-de-açúcar avaliada por meio de variáveis morfofisiológicas**. 65 f. Dissertação (Mestrado em Agricultura) - Unesp, Botucatu.

Pincelli RP, Silva MA (2012). Alterações morfológicas foliares em cultivares de cana-de-açúcar em resposta à deficiência hídrica. **Bioscience Journal** 28:546-556.

Piperidis N, Piperidis G, D'Hont A (2010) Molecular cytogenetics. In.: Henry RJ, Kole C (Eds.) (2010) **Genetics, genomics and breeding of sugarcane**. Boca Raton: CRC Press, p. 9-18.

Prasitsom C, Jubsab N, Klomsa-Ard P, Sriroth K, Keawsompong S (2019) Selection of SSR markers for drought resistant sugarcane in Thailand. **International Journal of Agricultural Technology** 15:997-1010.

Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. **Theoretical and Applied Genetics** 116:701-714.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. **Current opinion in plant biology** 5:94-100.

Ramesh P (2000) Effect of different levels of drought during the formative phase on growth parameters and its relationship with dry matter accumulation in sugarcane. **Journal of Agronomy and Crop Science** 185:83–89.

Ribaut JM, Pilet PE (1994) Water stress and indol-3yl-acetic acid content of maize roots. **Planta** 193:502-507.

Robert F, Pelletier J (2018) Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. **Frontiers in genetics** 9:1-11.

Rodrigues FA, Da Graça JP, De Laia ML, Nhani-Jr A, Galbiati JA, Ferro MIT, Ferro JA, Zingaretti SM (2011) Sugarcane genes differentially expressed during water deficit. **Biologia Plantarum** 55:43–53.

Rodrigues FA, De Laia ML, Zingaretti SM (2009) Analysis of gene expression profiles under water stress in tolerant and sensitive sugarcane plants. **Plant Science** 176:286-302.

Rodrigues JD, Jadoski CJ, Fagan EB, Ono EO, Soares LH, Neto DD (Eds.) (2018) Fisiologia da produção de cana-de-açúcar. São Paulo: Andrei, 176p.

Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M (2018) Splicing regulation by long noncoding RNAs. **Nucleic acids research** 46:2169-2184.

Santos JM, Duarte Filho LSC, Soriano ML, Silva PP, Nascimento VX, Barbosa GVS, Todaro AR, Ramalho Neto CE, Almeida C (2012) Genetic diversity of the main progenitors of sugarcane from the RIDESA germplasm bank using SSR markers. **Industrial Crops and Products** 40:145-150.

Santos RF, Carlesso R (1998) Déficit hídrico e os processos morfológico e fisiológico das plantas. **Revista Brasileira de Engenharia Agrícola e Ambiental** 2:287-294.

Schwartz SH, Qin X, Zeevaart JA (2003) Elucidation of the indirect pathway of abscisic acid biosynthesis by mutants, genes, and enzymes. **Plant physiology** 131:1591-1601.

Seok HY, Woo DH, Park HY, Lee SY, Tran HT, Lee EH, Nguyen LV, Moon YH (2016) AtC3H17, a non-tandem CCCH zinc finger protein, functions as a nuclear transcriptional activator and has pleiotropic effects on vegetative development, flowering and seed development in *Arabidopsis*. **Plant and Cell Physiology** 57:603-615.

Shan H et al. (2012) Heterologous expression of the chrysanthemum R2R3-MYB transcription factor CmMYB2 enhances drought and salinity tolerance, increases hypersensitivity to ABA and delays flowering in *Arabidopsis thaliana*. **Molecular biotechnology** 51:160-173.

Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. **Current opinion in plant biology** 3:217-223.

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. **Journal of experimental botany** 58:221-227.

Singh P, Suman A, Tiwari P, Arya N, Gaur A, Shrivastava AK (2008) Biological pretreatment of sugarcane trash for its conversion to fermentable sugars. **World Journal of Microbiology and Biotechnology** 24:667-673.

Singh S, Rao PG (1987) Varietal differences in growth characteristics in sugar cane. **The Journal of Agricultural Science** 108:245-247.

Smit MA, Singels A (2006) The response of sugarcane canopy development to water stress. **Field Crops Research** 98:91-97.

Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. **F1000Research**, 4:1-19.

Song Y, You J, Xiong L (2009) Characterization of OsIAA1 gene, a member of rice Aux/IAA family involved in auxin and brassinosteroid hormone responses and plant morphogenesis. **Plant molecular biology** 70:297-309.

Souza GM et al. (2019) Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. **GigaScience** 8:1-18.

Spannagl M, Nussbaumer T, Bader KC, Martis MM, Seidel M, Kugler KG, Gundlach H, Mayer KFX (2015) PGSB PlantsDB: updates to the database framework for comparative plant genome research. **Nucleic acids research** 44:1141-1147.

Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (Ed.) **Sugarcane improvement through breeding**. Amsterdam: Elsevier, p. 211-253.

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. **Bioinformatics** 23:1282-1288.

Taiz L, Zeiger E, Møller IM, Murphy A (2017) Fisiologia e desenvolvimento vegetal. Porto Alegre: Artmed, 888p.

Team R Core (2013) R: A language and environment for statistical computing. [S. l.].

Telles BR (2016) **Características fisiológicas e moleculares de uma cultivar de cana-de-açúcar tolerante à seca e submetida ao déficit hídrico prolongado**. 184 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Unesp, Jaboticabal.

Telles BR (2019) Prolonged water deficit reveals new profile of sugarcane gene expression and metabolic pathway related to tolerance. **Sugar Tech** 21:451-461.

Turchetto-Zolet AC, Turchetto C, Zanella CM, Passaia G (2017) Marcadores Moleculares na Era Genômica: Metodologias e Aplicações. Ribeirão Preto: Sociedade Brasileira de Genética, 180p.

Tyedmers J, Mogk A, Bukau B (2010) Cellular strategies for controlling protein aggregation. **Nature reviews Molecular cell biology** 11:777-788.

Udawat P, Jha RK, Sinha D, Mishra A, Jha B (2016) Overexpression of a cytosolic abiotic stress responsive universal stress protein (SbUSP) mitigates salt and osmotic stress in transgenic tobacco plants. **Frontiers in plant science** 7:1-21.

Unica - União da Indústria de Cana-de-Açúcar. Moagem de cana-de-açúcar e produção de açúcar e etanol - safra 2018/2019. Disponível em: <<https://www.unicadata.com.br/>> Acesso em: 27 set. 2019.

Unica - União da Indústria de Cana-de-Açúcar. Setor sucroenergético – Mapa de Produção. Disponível em: <<http://www.unica.com.br/mapa-da-producao/>> Acesso em: 12 ago. 2019.

Van Vliet AHM (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. **FEMS microbiology letters** 302:1-7.

VanBogelen RA, Hutton ME, Neidhardt FC (1990) Gene-protein database of *Escherichia coli* K-12: edition 3. **Electrophoresis** 11:1131-1166.

Vantini JS, Dedemo GC, Gimenez DJ, Fonseca LF, Tezza RI, Mutton MA, Ferro JA, Ferro MIT (2015) Differential gene expression in drought-tolerant sugarcane roots. **Genetics and Molecular Research** 14:7196-7207.

Vargas L et al. (2014) Drought tolerance conferred to sugarcane by association with *gluconacetobacter diazotrophicus*: a transcriptomic view of hormone pathways. **PLoS One** 9:1-37.

Vaseva I, Sabotič J, Šuštar-Vozlič J, Meglič V, Kidrič M, Demirevska K, Simova-Stoilova L (2012) The response of plants to drought stress: the role of dehydrins, chaperones, proteases and protease inhibitors in maintaining cellular protein function. In.: Neves DF, Sanz JD (Eds.) **Droughts: new research**. New York: Nova Science Publishers, p. 1-45.

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. **The Plant Journal** 45:523-539.

Vettore AL et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research** 13:2725–2735.

Vettore AL, Silva FRD, Kemper EL, Arruda P (2001) The libraries that made SUCEST. **Genetics and Molecular Biology** 24:1-7.

Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. **Trends in ecology & evolution** 11:158-162.

Wang D, Guo Y, Wu C, Yang G, Li Y, Zheng C (2008) Genome-wide analysis of CCCH zinc finger family in *Arabidopsis* and rice. **BMC genomics** 9:1-20.

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. **Planta** 218:1-14.

Wang Y, Mopper S, Hasenstein KH (2001) Effects of salinity on endogenous ABA, IAA, JA, and SA in *Iris hexagona*. **Journal of chemical ecology** 27:327-342.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews genetics** 10:57-63.

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R (2006) *Biologia Molecular do Gene*. Porto Alegre: Artmed, 728p.

Wei S et al. (2014) A rice calcium-dependent protein kinase OsCPK9 positively regulates drought stress tolerance and spikelet fertility. **BMC plant biology** 14:1-13.

Weller JI, Glick G, Ezra E, Zeron Y, Seroussi E, Ron M (2010) Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. **Animal genetics** 41:551-553.

White PJ, Broadley MR (2003) Calcium in plants. **Annals of botany** 92:487-511.

Whittaker RH (1970) The biochemical ecology of higher plants. In.: Sondheimer E, Simeone JB (Ed.) **Chemical Ecology**. New York: Academic Press, p.43-70.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic acids research** 18:6531-6535.

Xiong J (Ed.) (2006) Essential bioinformatics. San Francisco: Cambridge University Press, 339 p.

Xu J, Tian YS, Peng RH, Xiong AS, Zhu B, Jin XF, Gao F, Fu XY, Hou XL, Yao QH (2010) AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in *Arabidopsis*. **Planta** 231:1251-1260.

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. **Trends in plant science** 10:88-94.

Ye J et al. (2018) WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. **Nucleic acids research** 46:71-75.

Yordanov I, Velikova V, Tsonev T (2003) Plant responses to drought and stress tolerance. **Bulgarian Journal of Plant Physiology** special issue:187-206.

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. **European Patent Application** No 0534858.

Zhang J, Li Y, Jia HX, Li JB, Huang J, Lu MZ, Hu JJ (2015) The heat shock factor gene family in *Salix suchowensis*: a genome-wide survey and expression profiling during development and abiotic stresses. **Frontiers in Plant Science** 6:1-14.

Zhao D, Glaz B, Comstock JC (2010) Sugarcane response to water-deficit stress during early growth on organic and sand soils. **American journal of agricultural and biological sciences** 5:403-414.

Zhu JK (2001) Cell signaling under salt, water and cold stresses. **Current opinion in plant biology** 4:401-406.

Zhu JK (2002) Salt and drought stress signal transduction in plants. **Annual review of plant biology** 53:247-273.