

UNIVERSIDADE FEDERAL DE VIÇOSA

LAURA GONÇALVES COSTA MARTINS

**REGULAÇÃO DE GENES RESPONSIVOS A ESTRESSES ABIÓTICOS:
TOLERÂNCIA A ALUMÍNIO EM SORGO E DISTÚRBIO FISIOLÓGICO EM
EUCALIPTO**

**VIÇOSA - MINAS GERAIS
2020**

LAURA GONÇALVES COSTA MARTINS

**REGULAÇÃO DE GENES RESPONSIVOS A ESTRESSES ABIÓTICOS:
TOLERÂNCIA A ALUMÍNIO EM SORGO E DISTÚRBO FISIOLÓGICO EM
EUCALIPTO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção de título de *Doctor Scientiae*.

Orientadora: Elizabeth Pacheco Batista Fontes

Coorientadora: Christiane Eliza Motta Duarte

**VIÇOSA - MINAS GERAIS
2020**

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

M379r
2020
Martins, Laura Gonçalves Costa, 1992-
Regulação de genes responsivos a estresses abióticos :
tolerância a alumínio em sorgo e distúrbio fisiológico em
eucalipto / Laura Gonçalves Costa Martins. – Viçosa, MG, 2020.
102f. : il. (algumas color.) ; 29 cm.

Orientador: Elizabeth Pacheco Batista Fontes.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Stress (Fisiologia). 2. Sorgo. 3. Eucalipto. 4. Expressão
gênica. 5. Sequenciamento de nucleotídeo. I. Universidade
Federal de Viçosa. Departamento de Bioquímica e Biologia
Molecular. Doutorado em Genética e Melhoramento. II. Título.

CDD 22 ed. 571.2492

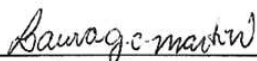
LAURA GONÇALVES COSTA MARTINS

**REGULAÇÃO DE GENES RESPONSIVOS A ESTRESSES ABIÓTICOS:
TOLERÂNCIA A ALUMÍNIO EM SORGO E DISTÚRBO FISIOLÓGICO EM
EUCALIPTO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção de título de *Doctor Scientiae*.

APROVADA: 20 de fevereiro de 2020.

Assentimento:



Laura Gonçalves Costa Martins
Autora



Elizabeth Pacheco Batista Fontes
Orientadora

AGRADECIMENTOS

Agradeço primeiramente a Deus, mais que ciência e a sabedoria, o criador delas, o responsável por todas as minhas conquistas e vitórias, meu criador e meu pai.

À minha família, minha mãe, meu padrasto, meus dois irmãos e, em especial a minha sobrinha Sofia, que me oferece alegria e amor todos os dias.

À Universidade Federal de Viçosa, por toda contribuição para meu aprendizado e formação.

À minha orientadora a contar do mestrado, Elizabeth Fontes, por toda sua competência, profissionalismo e excelência em orientar.

Ao meu Coorientador Jurandir, por todo seu conhecimento, solicitude e gentileza para com minha tese. Sou muito grata por tê-los como meus orientadores.

Ao Virgílio, presente constantemente ao meu lado, me dando apoio psicológico (ou não), e contribuindo com as diversas extrações de RNA.

À Christiane, por sua participação neste projeto, por toda parceria nos experimentos e pela amizade. Ao Ruan, por sua colaboração experimental. Aos bioinformatas, por toda ajuda nas análises de RNA-Seq e, a todos os membros atuais do LBMP.

Aos que não pertencem mais ao laboratório, agradeço à Iara, que sempre ajudou e esteve disposta a contribuir intelectualmente. À Maiana, minha base de formação acadêmica.

E por fim, aos secretários do programa de pós-graduação em Genética e Melhoramento, Marco Túlio e Odilon, pela gentileza em ajudar sempre que preciso.

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

RESUMO

MARTINS, Laura Gonçalves Costa, D.Sc., Universidade Federal de Viçosa, fevereiro de 2020. **Regulação de genes responsivos a estresses abióticos: tolerância a alumínio em sorgo e distúrbio fisiológico em eucalipto.** Orientadora: Elizabeth Pacheco Batista Fontes. Coorientadora: Christiane Eliza Motta Duarte.

As plantas estão sujeitas frequentemente a variações ambientais e para sobreviver às condições estressantes, elas utilizam mecanismos de defesa mediante alteração dos processos fisiológicos e moleculares. Esses estresses são em grande parte responsáveis pelo limiar entre potencial de produção de determinadas culturas e o seu real efeito de produtividade no campo. Nesta investigação, foram abordados dois temas de estresses abióticos em plantas, resistência a Al em sorgo e distúrbio fisiológico em eucalipto. Em sorgo, o gene *SbMATE* codifica um transportador transmembrana pertencente à família MATE (*Multidrug and Toxic Compound Extrusion family*) que é responsável pelo efluxo de citrato ativado por alumínio. Neste estudo, demonstrou-se que os transfatores *SbWRKY1* e *SbZNF1* são capazes de interagir com o promotor do gene *SbMATE* e promover a ativação transcricional do gene repórter *HIS3* em leveduras. Para ensaios de transativação *in planta*, diferentes fragmentos do promotor de *SbMATE* foram fusionados ao gene repórter *GUS*, utilizados para transformar *Arabidopsis thaliana*, pela técnica do mergulho floral e utilizadas para a obtenção de protoplastos. Experimentos complementares, utilizando regiões adicionais truncadas do promotor, possibilitaram o mapeamento do domínio *cis*-regulatório de ligação dos transfatores, o qual está contido na região deletada de 92pb, delimitada pelas posições -2102 a -2010. Análises por imunoprecipitação da cromatina (*ChIP*) reforçaram a ideia de que *SbWRKY1* e *SbZNF1* atuam no controle da expressão de *SbMATE* e possibilitaram a identificação de ligação direta dos transfatores ao promotor *SbMATE in vivo*. Embora, o mecanismo de tolerância a alumínio mediado pelo gene *SbMATE* tenha sido parcialmente esclarecido, a regulação transcricional desse gene ainda não foi elucidada por completo. Em relação ao capítulo II, o fenômeno do distúrbio fisiológico na espécie *E. grandis* tem sido recorrente em diversas regiões de plantio no sul e leste da Bahia, comprometendo sobremaneira a qualidade da madeira. Não se sabe exatamente a causa do aparecimento do “distúrbio fisiológico”, mas a ela tem sido atribuída a fatores climáticos e fatores genéticos. Os estágios iniciais do estabelecimento do distúrbio

fisiológico têm sido associados à perda de dominância apical e ramificação da parte inferior do caule ou da copa, com morte celular e necrose do caule primário, o que o diferencia de outras anomalias bióticas e abióticas. Os mecanismos moleculares envolvidos no desenvolvimento dos sintomas e na resposta de defesa da planta ao distúrbio são totalmente desconhecidos. Foi utilizada a tecnologia RNA-Seq para compreensão das bases moleculares deste mecanismo associados ao “distúrbio fisiológico”, a partir de amostras de caules do clone suscetível 520 e do clone resistente 1404. As amostras de caules dos referidos clones foram coletadas no campo de plantas crescidas no mesmo talhão sob as mesmas condições climáticas. Este delineamento experimental permitiu inferir que os genes diferencialmente expressos pelo contraste 520-S1 vs 520-S3 representavam o transcriptoma induzido pelo distúrbio fisiológico, enquanto que 520-S2 vs 520-S3 representava o transcriptoma induzido em recuperação do estado de estresse. O contraste 1404-R1 vs 520-S3 representa o transcriptoma induzido pelo somatório dos efeitos de resistência e de genótipo. O estabelecimento do distúrbio fisiológico nas amostras 520-S1 provocou uma reprogramação massiva na expressão gênica, resultando em 4286 genes diferencialmente expressos comparados com 520-S3. Entre os genes regulados positivamente houve um enriquecimento de genes na categoria morte celular, sistema imune inato, sistema antioxidativo, e metabolismo secundário ligados à resposta a estresses abióticos. Além disso, um estudo de metagenômica não identificou patógenos principais nas amostras analisadas. Estes resultados eliminam a possibilidade de que os sinais bióticos sejam o principal estímulo ao distúrbio fisiológico. Entre as categorias biológicas enriquecidas para os genes regulados negativamente, particularmente relevantes, foram a repressão extensiva em genes envolvidos na síntese de ácido graxo e no sistema de reparo do DNA. Em termos de diagnóstico molecular, repressão na síntese de ácidos graxos de cadeia longa pode explicar a maioria dos sintomas observados nos caules acometidos pelo distúrbio fisiológico, enquanto que repressão do sistema de reparo do DNA pode ser considerado como o disparo para ativação de vias de morte celular programada. Analisando o contraste 520-S2 vs 520-S3, apenas 1311 genes foram diferencialmente expressos, indicando um nível de estresse atenuado e estado de recuperação desses caules emergentes em relação à amostra suscetível assintomática S3. Com relação ao contraste 1404-R1 vs 520-S3, observou-se um enriquecimento de genes regulados positivamente na categoria tradução, com uma super-representação de genes de proteínas

ribossomais e da maquinaria de síntese de proteínas, e na categoria sistema de reparo do DNA. Enquanto a indução massiva de genes de proteínas ribossomais deve refletir diferenças em genótipos, já que não se observaram mudanças significativas no contraste 520-S1 vs 520-S3, a regulação positiva de genes do sistema de reparo do DNA pode representar, pelo menos em parte, o efeito da resistência. Entre os genes *down* regulados, destacam-se genes envolvidos no sistema inato de plantas, tanto ETI como PTI. Além desses, um subconjunto de genes de resistência do tipo TIR-NBS-LRR foram extensivamente *down* regulados, substanciando o argumento de que sinais bióticos não correspondem ao principal ativador do distúrbio fisiológico. Enquanto a análise de expressão por qRT-PCR de um conjunto selecionado de genes diferencialmente expressos confirmou os resultados das análises de RNA-Seq, não foi possível demonstrar que folhas sintomáticas mimetizam o transcriptoma induzido pelo distúrbio fisiológico determinado em caules sintomáticos. Estudos posteriores com base em análises de possíveis SNPs (*single nucleotide polymorphism*) associados ao distúrbio, deverão ser realizadas para definição de marcadores moleculares ligados ao distúrbio fisiológico.

Palavras-chave: Estresse abiótico. Sorgo. Eucalipto. Expressão gênica. RNA-Seq.

ABSTRACT

MARTINS, Laura Gonçalves Costa, D Sc., Universidade Federal de Viçosa, February, 2020. **Regulation of genes responsive to abiotic stresses: aluminum tolerance in sorghum and physiological disturbance in eucalyptus.** Adviser: Elizabeth Pacheco Batista Fontes. Co-adviser: Christiane Eliza Motta Duarte.

Plants are often subject to environmental variations and to survive stressful conditions, they use defense mechanisms by altering physiological and molecular processes, in order to maintain their growth and reproduction. These stresses are largely responsible for the threshold between the production potential of certain crops and their real productivity effect in the field. In this investigation, two themes of abiotic stresses in plants were addressed, resistance to Al in sorghum and physiological disturbance in eucalyptus. In sorghum, the *SbMATE* gene encodes a transmembrane transporter belonging to the MATE family (Multidrug and Toxic Compound Extrusion family) which is responsible for the aluminum-activated citrate efflux. In this study, it was demonstrated that the SbWRKY1 and SbZNF1 transactors are capable of interacting with the *SbMATE* gene promoter and promoting the transcriptional activation of the HIS3 reporter gene in yeasts. For in-plant transactivation assays, different fragments of the *SbMATE* promoter were fused to the GUS reporter gene, used to transform *Arabidopsis thaliana*, by the technique of floral dipping and used to obtain protoplasts. Complementary experiments, using additional truncated regions of the promoter, enabled the mapping of the cis-regulatory domain of binding of the transactors, which is contained in the deleted region of 92bp, delimited by the positions -2102 to -2010. Chromatin immunoprecipitation (ChIP) analyzes reinforced the idea that SbWRKY1 and SbZNF1 act in the control of *SbMATE* expression and made it possible to identify the direct link between transactors and the *SbMATE* promoter in vivo. Although the mechanism of aluminum tolerance mediated by the *SbMATE* gene has been partially clarified, the transcriptional regulation of this gene has not yet been fully elucidated. Concerning Chapter II, the phenomenon of physiological disturbance in the species *E. grandis* has been recurrent in several planting regions in the south and east of Bahia, greatly compromising the quality of the wood. It is not known exactly the cause of the appearance of the “physiological disturbance”, but it has been attributed to climatic factors and genetic factors. The initial stages of the establishment of the physiological disorder have been associated with loss of

apical dominance and branching of the lower part of the stem or canopy, with cell death and necrosis of the primary stem, which differentiates it from other biotic and abiotic anomalies. The molecular mechanisms involved in the development of symptoms and in the plant's defense response to the disorder are totally unknown. RNA-Seq technology was used to understand the molecular bases of this mechanism associated with “physiological disturbance”, from stem samples from susceptible clone 520 and resistant clone 1404. The stem samples from the said clones were collected in the plant field grown in the same plot under the same climatic conditions. This experimental design made it possible to infer that the genes differentially expressed by the 520-S1 vs 520-S3 contrast represented the transcriptome induced by the physiological disorder, while 520-S2 vs 520-S3 represented the transcriptome induced in recovery from the state of stress. The contrast 1404-R1 vs 520-S3 represents the transcriptome induced by the sum of the effects of resistance and genotype. The establishment of the physiological disturbance in the 520-S1 samples caused a massive reprogramming of gene expression, resulting in 4286 differentially expressed genes compared to 520-S3. Among the up-regulated genes, there was an enrichment of genes in the category cell death, innate immune system, antioxidative system, and secondary metabolism linked to the response to abiotic stresses. In addition, a metagenomics study did not identify major pathogens in the analyzed samples. These results eliminate the possibility that the biotic signals are the main stimulus to the physiological disorder. Among the biological categories enriched for the negatively regulated genes, particularly relevant, were the extensive repression in genes involved in the fatty acid synthesis and the DNA repair system. In terms of molecular diagnosis, repression in the synthesis of long-chain fatty acids can explain most of the symptoms observed in the stems affected by the physiological disorder, while repression of the DNA repair system can be considered as the trigger for activation of death pathways. Analyzing the 520-S2 vs 520-S3 contrast, only 1311 genes were differentially expressed, indicating an attenuated stress level and recovery status of these emerging stems in relation to the asymptomatic susceptible sample S3. Regarding contrast 1404-R1 vs 520-S3, there was an enrichment of positively regulated genes in the translation category, with an over-representation of ribosomal protein genes and protein synthesis machinery, and in the DNA repair system category. While the massive induction of ribosomal protein genes should reflect differences in genotypes, since

no significant changes were observed in the 520-S1 vs 520-S3 contrast, positive regulation of genes in the DNA repair system may represent, at least in part, the effect of resistance. Among the down-regulated genes, there are genes involved in the innate plant system, both ETI and PTI. In addition to these, a subset of TIR-NBS-LRR-like resistance genes has been extensively down-regulated, substantiating the argument that biotic signals do not correspond to the main activator of the physiological disorder. While qRT-PCR expression analysis of a selected set of differentially expressed genes confirmed the results of RNA-Seq analyzes, it was not possible to demonstrate that symptomatic leaves mimic the transcriptome induced by the physiological disorder determined in symptomatic stems. Further studies based on analyzes of possible SNPs (single nucleotide polymorphism) associated with the disorder, should be carried out to define molecular markers linked to the physiological disorder.

Keywords: Abiotic stress. Sorghum. Eucalyptus. Gene expression. RNA-Seq.

SUMÁRIO

INTRODUÇÃO GERAL	11
REFERÊNCIAS BIBLIOGRÁFICAS.....	16
CAPÍTULO I.....	19
CAPÍTULO II.....	35
RESUMO	36
1. INTRODUÇÃO.....	38
2. MATERIAL E MÉTODOS.....	44
2.1 Coleta do Material vegetal.....	44
2.2 Extração de RNA total das amostras de eucalipto.....	44
2.3 Integridade e preparo das amostras de RNA para envio ao sequenciamento.....	45
2.4 Sequenciamento de alto desempenho do transcriptoma – RNA (Seq).....	46
2.4.1 Construção da Biblioteca de cDNA.....	46
2.4.2 Formação de clusters paired-end.....	47
2.5 Bioinformática.....	49
2.5.1 Verificação da qualidade das leituras geradas.....	50
2.5.2 Alinhamento das leituras, montagem de transcritos, análises de expressão diferencial dos genes e mapeamento das redes de co-expressão e modelos de sinalização celular alterados.....	52
2.6 Extração de RNA dos clones de eucalipto e síntese de cDNA.....	53
2.7 Análises de transcritos por qRT-PCR.....	53
3. RESULTADOS.....	56
3.1 Qualidade das amostras selecionadas para o sequenciamento.....	56
3.2 Avaliação de qualidade de leitura geradas pelo sequenciamento, com FastQC, TRIMAGEM e limpeza das sequências de Reads.....	57
3.3 Transcriptoma induzido pelo distúrbio fisiológico.....	58
3.4 Análise de expressão gênica para os genes candidatos.....	85
4. DISCUSSÃO.....	90
5. REFERÊNCIAS.....	95

INTRODUÇÃO GERAL

As plantas estão constantemente expostas a condições ambientais adversas. Dentre os principais grupos de estresses abióticos que limitam produtividade vegetal, destacam-se o déficit hídrico, estresse nutricional, caracterizado pela relação de absorção de nutrientes das plantas com as propriedades químico-físicas dos solos - pH, mineralogia, presença de toxidez, metais pesados e salinidade, além de estresse causado pela temperatura, caracterizada pela variação térmica (Souza & Ribeiro, 2015). As plantas apresentam sistemas biológicos complexos abrangendo milhares de genes, proteínas, moléculas reguladoras, agentes de sinalização e compostos químicos diferentes, que estabelecem centenas de rotas e redes interligadas. Sob condições de estresse, diferentes rotas bioquímicas e redes de sinalização ajustam-se homeostaticamente para minimizar os impactos negativos do estresse e manter o equilíbrio metabólico (Taiz et al., 2017).

Merecem destaques como maiores restrições à produtividade agrícola, os solos ácidos associados a metais tóxicos, uma vez que afetam 50% das terras aráveis, sendo especialmente comum em regiões tropicais e subtropicais devido à maior pluviosidade (Von Uexkull and Mutert, 1995). Entre os metais, o Al, caracterizado por ser o metal mais abundante e o terceiro elemento mais comum na crosta terrestre, destaca-se como um dos maiores problemas em solos com $\text{pH} \leq 5,0$. Em soluções ácidas ($\text{pH} < 5,0$), o Al solubiliza-se passando à forma de octaedro hexahidratado $[\text{Al}(\text{H}_2\text{O})_6]^{3+}$, também conhecido por Al^{3+} ; trata-se de um elemento anfótero que pode atuar como cátion em meio ácido e como ânion em meio básico, sendo o pH o principal fator de controle de sua disponibilidade no solo (Malavolta, 1980). O aumento do pH causa desprotonações sucessivas formando $\text{Al}(\text{OH})_2^+$ e $\text{Al}(\text{OH})^{+2}$. Em pH neutro, o alumínio apresenta-se na forma de $\text{Al}(\text{OH})_3$ (gibbsite), composto que não é tóxico às raízes das plantas, sendo esta a base química da neutralização do efeito fitotóxico do Al^{3+} (Magalhães, 2007). À medida que o pH aumenta para os valores comumente encontrados no citoplasma, ($\text{pH}=7,4$), o íon aluminato $\text{Al}(\text{OH})_4^-$ domina a especiação (Kochian et al., 1995).

O processo natural de acidificação do solo é muitas vezes intensificado por práticas agrícolas, a lixiviação retira elementos químicos do solo, em especial cálcio e magnésio através da água de percolação, que substitui as bases por hidrogênio e alumínio,

intensificando assim, a acidificação (Malavolta, 1980; Fageria, 1989; Maria et al., 1993; Fernandes, 2006). Tais problemas podem ser resolvidos pela etapa de calagem, num processo que se aplica calcário com o objetivo de elevar teores de cálcio e magnésio, neutralizando os íons H^+ e Al^{3+} e corrigindo o pH do solo, para um desenvolvimento satisfatório das culturas (Lopes et al., 1991).

Espécies Al^{3+} são fitotóxicas para as plantas, mesmo em concentrações micromolares e são responsáveis por afetar o crescimento e o desenvolvimento das raízes (Delhaize et al., 2012; Sade et al., 2016), além de diminuir a absorção de nutrientes, o que é desfavorável para o desenvolvimento de plantas sensíveis a esse elemento. A carga elétrica Al^{3+} permite que o Al reaja com numerosos sítios nas células, alvos potenciais de injúria, incluindo a parede celular, citoesqueleto e membrana plasmática (Kochian, 2004). Na parede celular, o Al pode ligar-se aos grupos carboxílicos das pectinas e das proteínas diminuindo a extensibilidade (Delhaize, 1995). A membrana plasmática parece ser o alvo primário dos danos provocados pelo Al (Wagatsuma, 1995), possui a superfície carregada negativamente, representando um alvo sensível à fitotoxicidade do Al (Ma, 2007. Panda et al., 2009). O Al liga-se fortemente aos fosfolipídios, modificando sua composição (Jones, 1997; Delhaize, 1995; Peixoto et al., 2001; Khan et al., 2009) reduzindo assim, a fluidez das membranas e aumentando a densidade de empacotamento dos lipídios (Chen et al., 1991). Uma vez no citoplasma celular, pode interromper a dinâmica do citoesqueleto, interagindo com os microtúbulos e filamentos de actina (Kochian, 2004), além de poder afetar as funções da respiração mitocondrial (Panda, 2008) e induzir o estresse oxidativo (Richards, 1998).

O Al^{3+} caracteriza-se como principal impedimento na produção, reduzindo o rendimento das culturas em solos nesta condição (Delhaize & Ryan, 1995) e inibindo o crescimento radicular, resultando em um sistema radicular pouco desenvolvido e incapaz de explorar camadas mais profundas de solo (Kochian, 1995; Sade et al., 2016). O alumínio, também provoca a diminuição da permeabilidade da membrana plasmática (Ahn, 2001), por meio da interação com fosfolipídios de membrana (Jones, 1997), peroxidação de lipídeos (Peixoto, 2001; Yamoto, 2001), além da disfunção mitocondrial, aumento de espécies reativas do oxigênio (ROS) (Yamamoto, 2002), além da deposição de calose e danos ao DNA (Sade et al., 2016).

Apesar da toxidez causada por Al^{3+} , algumas espécies possuem a capacidade de tolerar altas concentrações desse metal, devido a sua complexação com ácidos orgânicos exsudados pelo sistema radicular e ao genótipo vegetal que pode promover a capacidade de adequar às condições físico-químicas adversas, minimizando os problemas causados pela baixa produtividade em solos ácidos (Freitas et al., 2006; Nolla et al., 2007). O melhoramento genético para tolerância ao alumínio é considerado uma alternativa viável para aumentar o desempenho de plantas economicamente importantes cultivadas em solos ácidos ricos nesse metal. Deste modo, muitos laboratórios ao redor do mundo têm concentrado seus esforços em identificar e caracterizar a tolerância das plantas a níveis tóxicos de Al em solos ácidos e na busca por genótipos tolerantes (Ma, 2001; Panda, 2009).

Em sorgo, a herança da tolerância ao Al foi estudada em uma população derivada do cruzamento das linhagens de sorgo SC283 (tolerante) e BR007 (sensível). Os resultados mostraram tratar-se de uma herança monogênica, sendo o gene de tolerância ao Al, localizado no loco *Alt_{SB}* (Magalhães et al., 2004). Posteriormente, o gene *SbMATE* foi isolado pela técnica de clonagem posicional (Magalhães et al., 2007). Estes autores mostraram que o gene *SbMATE* possui maior expressão no ápice radicular e codifica um transportador de membrana pertencente à família multigênica MATE. Na linhagem tolerante, os níveis de expressão do gene *SbMATE* aumentam ao longo do tempo de exposição ao Al. O citrato liberado em resposta ao Al forma um complexo estável e insolúvel com os íons Al (atóxico), prevenindo assim a sua absorção pela planta (Magalhães et al., 2007). O gene *SbMATE* codifica um transportador transmembrana pertencente à família MATE (*Multidrug and Toxic Compound Extrusion family*) que é responsável pelo efluxo de citrato ativado por alumínio. A identificação de elementos *cis*-regulatórios no promotor do gene *SbMATE* e de fatores transcricionais que controlam a expressão do referido gene, constituem etapas fundamentais para o entendimento do mecanismo de resistência à Al mediado por *SbMATE*. Desta forma, estudos com base na regulação da expressão deste gene são relevantes como estratégia para desenvolvimento de cultivares de sorgo tolerantes ao Al.

Um problema adicional no Brasil, provavelmente resultante de mudanças em condições climáticas em áreas de plantio de eucalipto, refere-se ao “distúrbio fisiológico”, que se caracteriza por um conjunto de sintomas que resultam em perda de dominância apical associado à necrose do caule principal, que afetam diretamente a qualidade da madeira e

geram prejuízo para a cultura. O conjunto de sintomas responsáveis pelo distúrbio diferencia de outras anomalias causadas por fatores bióticos e abióticos. A incidência deste fenômeno no eucalipto tem sido recorrente em algumas regiões do Brasil.

A perda de dominância apical é responsável pela bifurcação em eucalipto e pode ser causada por uma série de fatores, dentre eles, os genéticos, ambientais e pela interação dos fatores genéticos com os ambientais. Nos casos onde ocorre essa interação, são estimados a herdabilidade do carácter em questão.

O denominador comum de todos os sintomas, parece ser a ativação descontrolada de vias de sinalização de morte celular programada (PCD; *programmed cell death*). PCD é uma denominação geral referente à morte geneticamente regulada de células individuais. É um aspecto essencial do desenvolvimento normal da planta, mas também pode ser induzida em resposta ao estresse (Taiz et al., 2007). A morte celular programada é responsável por uma sequência de eventos que levam a destruição controlada e organizada das células, o conjunto de mecanismos com características morfológicas são bem definidos: degradação nuclear e condensação da cromatina seguida da perda do DNA, degradação dos cloroplastos e o rompimento do vacúolo o qual libera enzimas (nucleases e proteases) que causam acidificação do meio, contribuindo para o rompimento do núcleo e do vacúolo (Gunawardena et al., 2004). Embora este fenômeno afete consideravelmente a produção da cultura, os agentes abióticos indutores, bem como os genes da planta envolvidos nesta resposta ainda não foram elucidados.

Consequentemente, ao contrário de mecanismos de toxidez e de resistência de plantas ao Al, cujas bases moleculares são mais bem compreendidas, o “distúrbio fisiológico” em eucalipto é um fenômeno biológico, cujas bases moleculares e indução do estado de estresse são absolutamente desconhecidos. Assim sendo, os objetivos desse trabalho se concentraram em dois focos principais de profundidade científica diferenciada. No primeiro capítulo, intitulado “*Repeat variants for the SbMATE transporter protect sorghum roots from aluminum toxicity by transcriptional interplay in cis and trans*”, foram abordados aspectos moleculares de controle da expressão do gene *SbMATE* que confere resistência ao Al em sorgo, resultando em avanços significativos de conhecimento na área molecular. Por outro lado, os objetivos do segundo capítulo se restringiram em analisar a variação global de expressão gênica induzidas pelo estado de distúrbio fisiológico em eucalipto além da

identificação de genes diferencialmente expressos em genótipos contrastantes para resistência e suscetibilidade ao estado de estresse. Os resultados permitiram diagnosticar molecularmente o estado de distúrbio fisiológico.

REFRÊNCIAS BIBLIOGRÁFICAS

Ahn S J, Sivaguru M, Osawa H, Chung GC, Matsumoto H (2001). Aluminum inhibits the H⁺-ATPase activity by permanently altering the plasma membrane surface potentials in squash roots. *Plant Physiology*, **126**, 1381-1390

Chen J, Sucoff E I, Stadelmann E J (1991). Aluminum and temperature alteration of cell membrane permeability of *Quercus rubra*. *Plant Physiol*, **96**:644-649.

Delhaize E, Ryan P R (1995). Aluminum toxicity and tolerance in plants. *Plant Physiology*, **107**, 315-321.

Delhaize, E.; Ma, J.F.; Ryan, P.R. (2012) **Transcriptional regulation of aluminium tolerance genes** Trends in Plant Science June 2012, Vol. 17, No. 6.

Fageria, N. K. Solos Tropicais e Aspectos Fisiológicos das Culturas. 425p. Brasília: EMBRAPA – DPU, 1989.

Fernandes, M. S. (2006). Nutrição mineral de plantas, 432p. Viçosa: SBCS.

Freitas, F. A.; Kopp, M. M.; Sousa, R. O.; Zimmer, P. D.; Carvalho, F. I. F.; Oliveira, A. C. (2006). Absorção de P, Mg, Ca e K e tolerância de genótipos de arroz submetidos a estresse por alumínio em sistemas hidropônicos. *Revista Ciência Rural*, Santa Maria, v. 36, n. 1, p. 72-79.

Gunawardena, A. H. L. A. N.; Greenwood, J. S.; Gengler, N. G. (2004). Programmed Cell Death Remodels Laccase Plant Leaf Shape during Development. *The plant cell*. Canada, v. 16, p. 60-73.

Jones D L, Kochian L V (1997). Aluminum interaction with plasma membrane lipids and enzyme metal binding sites and its potential role in Al cytotoxicity. *FEBS Letters* **400** (1997) 51-57.

Khan M S H, Tawaraya K, Sekimoto H, Koyama H, Kobayashi Y, Murayama T, Chuba T, Kambayashi M, Shiono Y, Uemura M, Ishikawa S, Wagatsuma T (2009). Relative abundance of Δ^5 -sterols in plasma membrane lipids of root-tip cells correlates with aluminum tolerance of rice. *Physiologia Plantarum*, **135**, 73-83.

Kochian L V (1995). Mechanisms of aluminium toxicity and resistance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**:237-60.

Kochian L V, Hoekenga O A, Pineros M A (2004). How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. *Annual Review Plant Biology*, **55**, 459-493.

Lopes, A.S.; Silva, M. de C.; Guilherme, L.R.G. (1991). Acidez do solo e calagem. 3.ed. São Paulo: Associação Nacional para Difusão de Adubos e Corretivos Agrícolas, 15p. (ANDA. Boletim técnico, 1).

Ma J F, Ryan P R, Delhaize E (2001) Aluminium tolerance in plants and the complexing role of organic acids. *Trends in Plant Science*, **6**, 273-278.

Ma J F (2007) Syndrome of aluminum toxicity and diversity of aluminum resistance in higher plants. *International Review of Cytology*, **264**, 225-253.

Malavolta, E. (1980). Elementos de Nutrição Mineral de Plantas. São Paulo: Editora agrônômica Ceres Ltda. 251 p.

Magalhães J V, Garvin D F, Wang Y, Sorrells M E, Klein P E, Schaffert R E, Li L, Kochian L V (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. *Genetics* **167**: 1905-1914.

Magalhães J V, Liu J, Guimarães C T, Lana U G P, Alves V M C, Wang Y-H, Schaffert R E, Hoekenga O A, Piñeros M A, Shaff J E, Klein P E, Carneiro N P., Coelho C M, Trick H N, Kochian L V (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nature Genetics*, **39**, 1156-1161.

Maria, I. C.; Rossetto, R.; Ambrosano, E. J.; Castro, O. M. (1993). Efeito da adição de diferentes fontes de cálcio no movimento de cátions em colunas de solo. *Revista Scientia Agrícola*, Piracicaba, v. 50, n. 1, p. 87-98.

Nolla, A.; Schlindwein, J. A.; Anghinoni, I. Crescimento, morfologia radicular e liberação de compostos orgânicos por plântulas de soja em função da atividade de alumínio na solução do solo de campo natural. *Revista Ciência Rural*, Santa Maria, v. 37, n.1, p. 97-101, 2007.

Panda S K, Yamamoto Y, Kondo H, Matsumoto H. (2008) Mitochondrial alterations related to programmed cell death in tobacco cells under aluminium stress. *C. R. Biologies* **331** 597-610

Panda S K, Baluska F, Matsumoto H (2009) Aluminum stress signaling in plants. *Plant Signaling & Behavior*, **4**, 592-597.

Peixoto P H P, Cambraia J, Sant'anna R, Mosquim P R, Moreira M A (2001) Aluminum effects on fatty acid composition and lipid peroxidation of a purified plasma membrane fraction of root apices of two sorghum cultivars. *Journal of Plant Nutrition*, **24**, 1061-1070.

Richards, K.D.; Schott, E.J.; Sharma, Y.K.; Davis, K.R. and Gardner, R.C. (1998) Aluminum Induces Oxidative Stress Genes in *Arabidopsis thaliana* *Plant Physiol.* (1998) **116**: 409-418.

Sade H, Meriga B, Surapu V, Gadi J, Sunita M. S. L, Suravajhala P, Kavi Kishor PB (2016) Toxicity and tolerance of aluminum in plants: tailoring plants to suit to acid soils *Biometals* **29**:187–210.

Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2017). *Fisiologia e desenvolvimento vegetal*. Artmed Editora.

Von Uexkull H R, Mutert E (1995). Global extent, development and economic impact of acid soils. *Plant and Soil*, Dordrecht, **v.171**, p.1-15.

Wagatsuma T, Ishikawa S, Obata H, Tawaraya K, Katoh A, TOHDA S. (1995) Plasma membrane of younger and outer cells in the primary specific site for aluminium toxicity in roots. *Plant Soil* **171**: 105-112.

Yamamoto Y, Kobayashi Y, Davi S R, Rikiishi S, Matsumoto H (2002). Aluminium toxicity is associated with mitochondrial dysfunction and the production of reactive oxygen species in plant cells. *Plant Physiology* vol. **128**:63-72.

Yamamoto Y, Kobayashi Y, Matsumoto H (2001) Lipid peroxidation is an early symptom triggered by aluminium, but not primary cause of elongation inhibition in pea roots. *Plant Physiology* v. **125**: 199-208.

CAPÍTULO I

Repeat variants for the *SbMATE* transporter protect sorghum roots from aluminum toxicity by transcriptional interplay in *cis* and *trans*

Artigo publicado

Janaina O. Melo, Laura G.C. Martins, Beatriz A. Barros, Maiana R. Pimenta, Ubiraci G.P. Lana, Christiane E. M. Duarte, Maria M. Pastina, Claudia T. Guimaraes¹, Robert E. Schaffert¹, Leon V. Kochian, Elizabeth P.B. Fontes, Jurandir V. Magalhaes (2019). Repeat variants for the *SbMATE* transporter protect sorghum roots from aluminum toxicity by transcriptional interplay in *cis* and *trans*. (doi:10.5061/dryad.18p3h04).

Páginas 20 a 34



Repeat variants for the *SbMATE* transporter protect sorghum roots from aluminum toxicity by transcriptional interplay in *cis* and *trans*

Janaina O. Melo^{a,b,1,2}, Laura G. C. Martins^{c,1}, Beatriz A. Barros^a, Maiana R. Pimenta^{c,3}, Ubiraci G. P. Lana^a, Christiane E. M. Duarte^c, Maria M. Pastina^a, Claudia T. Guimaraes^a, Robert E. Schaffert^a, Leon V. Kochian^d, Elizabeth P. B. Fontes^{c,4}, and Jurandir V. Magalhaes^{a,b,4}

^aEmbrapa Maize and Sorghum, Brazilian Agricultural Research Corporation, 35701-970 Sete Lagoas, MG, Brazil; ^bDepartamento de Biologia Geral, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil; ^cDepartment of Biochemistry and Molecular Biology, BIOAGRO, Universidade Federal de Viçosa, 36570-000 Viçosa, MG, Brazil; and ^dGlobal Institute for Food Security, University of Saskatchewan, SK S7N 4J8 Saskatoon, Canada

Edited by Julia Bailey-Serres, University of California, Riverside, CA, and approved November 14, 2018 (received for review May 16, 2018)

Acidic soils, where aluminum (Al) toxicity is a major agricultural constraint, are globally widespread and are prevalent in developing countries. In sorghum, the root citrate transporter *SbMATE* confers Al tolerance by protecting root apices from toxic Al³⁺, but can exhibit reduced expression when introgressed into different lines. We show that allele-specific *SbMATE* transactivation occurs and is caused by factors located away from *SbMATE*. Using expression-QTL mapping and expression genome-wide association mapping, we establish that *SbMATE* transcription is controlled in a bipartite fashion, primarily in *cis* but also in *trans*. Multiallelic promoter transactivation and ChIP analyses demonstrated that intermolecular effects on *SbMATE* expression arise from a WRKY and a zinc finger-DHHC transcription factor (TF) that bind to and *trans*-activate the *SbMATE* promoter. A haplotype analysis in sorghum RILs indicates that the TFs influence *SbMATE* expression and Al tolerance. Variation in *SbMATE* expression likely results from changes in tandemly repeated *cis* sequences flanking a transposable element (a miniature inverted repeat transposable element) insertion in the *SbMATE* promoter, which are recognized by the Al³⁺-responsive TFs. According to our model, repeat expansion in Al-tolerant genotypes increases TF recruitment and, hence, *SbMATE* expression, which is, in turn, lower in Al-sensitive genetic backgrounds as a result of lower TF expression and fewer binding sites. We thus show that even dominant *cis* regulation of an agronomically important gene can be subjected to precise intermolecular fine-tuning. These concerted *cis/trans* interactions, which allow the plant to sense and respond to environmental cues, such as Al³⁺ toxicity, can now be used to increase yields and food security on acidic soils.

transcriptional regulation | abiotic stress | transporters | expression QTL
| MITE transposon

Decisions in plant breeding often reflect the complex interplay between noncoding DNA sequences, acting locally in chromatin, and *trans*-regulatory elements driving gene expression via intermolecular interactions. Studies in *Drosophila* have shown that *cis* elements are evolutionarily important (1), whereas *trans* factors play pivotal roles in regulating plant stress responses in *Arabidopsis* (2). However, the manner in which *cis* and *trans* factors interact to control phenotypic expression is less clear.

Half the world's agricultural soils are highly acidic (3), which solubilizes Al³⁺ into the soil solution, damaging plant roots and reducing yields. The Al-activated root citrate transporter *SbMATE*, which underlies Al tolerance via formation of nontoxic Al-citrate complexes in the rhizosphere (4), increased grain yield by 0.6 ton·ha⁻¹ for sorghum grown on acidic soil (5). *SbMATE* SNPs were associated to sorghum grain yield production in West Africa (6), where sorghum is a staple food. This makes *SbMATE* important for global food security.

SbMATE expression is up-regulated by Al³⁺ in a time-dependent fashion (4) and is highly correlated with Al tolerance (7). A Tourist-

like miniature inverted repeat transposable element (MITE) (8) and its flanking sequences, which are repeated in tandem, were found 2 kb upstream of *SbMATE*. Variation in the number of these tandem repeats in different sorghum lines was positively correlated with Al tolerance. Nevertheless, introgression of the *Al_{TSB}* locus, where *SbMATE* resides (9), into Al-sensitive recurrent parents resulted in reduced *SbMATE* expression and Al tolerance; this suggested involvement of accessory loci acting in *trans* (7). The current study focuses on the elucidation of the role of the MITE repeats in *SbMATE* transcriptional regulation, and on the dissection of the genetic background effects that can reduce *SbMATE* expression.

Significance

Aluminum (Al³⁺) on acidic soils, which represent half of the world's agricultural lands, damages plant roots. In Africa, where sorghum is a staple food, 20% of the agricultural soils are acidic, significantly reducing yields. *SbMATE* confers sorghum Al tolerance via root citrate exudation into the soil, where citrate binds and detoxifies Al³⁺, but shows reduced expression in some genetic backgrounds. This phenomenon results from the action of a variable tandem repeat flanking a transposon in the *SbMATE* promoter working in concert with WRKY and zinc finger-DHHC proteins, which bind to the *SbMATE* promoter and regulate expression in response to Al³⁺. We can now select for superior alleles of these transcription factors to maximize *SbMATE* expression, thereby contributing to global food security.

Author contributions: J.V.M. designed and supervised research; J.O.M., L.G.C.M., B.A.B., M.R.P., U.G.P.L., C.E.M.D., and R.E.S. performed research; M.M.P. contributed new reagents/analytic tools; J.O.M., L.G.C.M., B.A.B., M.R.P., U.G.P.L., C.E.M.D., M.M.P., C.T.G., L.V.K., and J.V.M. analyzed data; J.V.M. wrote the paper; R.E.S. assisted with genetic stock design and construction; and E.P.B.F. conceived and supervised TF validation in yeast and *Arabidopsis*.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: Data associated with this paper are available to download from the Dryad Digital Repository (doi:10.5061/dryad.18p3h04). The uploaded data (December 10, 2018) include SNP physical positions and association *P* values with Al tolerance and *SbMATE* expression.

¹J.O.M. and L.G.C.M. contributed equally to this work.

²Present address: Departamento de Ciências Básicas (DCB), Universidade Federal dos Vales do Jequitinhonha e Mucuri, 39100-000 Diamantina, MG, Brazil.

³Present address: Núcleo de Graduação em Agronomia, Universidade Federal do Sergipe, 49680-000 Nossa Senhora da Glória, SE, Brazil.

⁴To whom correspondence may be addressed. Email: bbfontes@ufv.br or jurandir.magalhaes@embrapa.br.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808400115/-/DCSupplemental.

Published online December 13, 2018.

Results

***SbMATE* Expression Is Influenced by the Genetic Background.** We studied global (i.e., joint expression of *SbMATE* alleles) and allele-specific expression of *SbMATE* (Fig. 1) to distinguish between *cis* and *trans* regulatory effects. This was done by using stocks derived from the low *SbMATE*-expressing Al-sensitive line BR012 crossed with SC566, a very Al-tolerant line with high *SbMATE* expression (7). We generated a homozygous stock in which the *SbMATE* allele from SC566 was introgressed into the BR012 genetic background [SC566-near isogenic line (NIL)]. BR012 × SC566 and BR012 × SC566-NIL both have *SbMATE* in heterozygosity, but they have different genetic backgrounds. Although BR012 × SC566 has a hybrid background, BR012 × SC566-NIL has the homozygous background of BR012. Compared with SC566, there was a consistent reduction of global *SbMATE* expression (Fig. 1A) in the SC566-NIL. *SbMATE* expression was higher in the hybrid background (BR012 × SC566) than in the BR012 background in the BR012 × SC566-NIL. Therefore, *SbMATE* expression is reduced in the Al-sensitive background.

SbMATE allele-specific expression was quantified relative to expression in the Al-tolerant and Al-sensitive parents, SC566 and BR012 (Fig. 1B and C). Here, allele-specific expression was based on a T/A single nucleotide polymorphism (SNP) in the first exon of *SbMATE*, with the A allele present in SC566 and the T allele present in BR012 (7). A marked, 210-fold up-regulation of the Al-sensitive allele (T) was observed in the BR012 × SC566 hybrid (Fig. 1B). Expression of the Al-sensitive allele of *SbMATE* was greatly reduced when present in the BR012 genome (BR012 × SC566-NIL) compared with the hybrid genome (BR012 × SC566; Fig. 1B). Expression changes for the Al-tolerant allele (A) were relatively subtle (Fig. 1C), but expression was reduced in all stocks harboring the BR012 genome compared with SC566.

These findings indicate that *SbMATE* expression is influenced by *trans*-acting factors whose favorable alleles are donated by the Al-tolerant line SC566. These factors are unlinked to *AltsB*, as allele-specific expression was higher in hybrids (BR012 × SC566) compared with the NIL hybrid (BR012 × SC566-NIL), where the *AltsB* locus is heterozygous but within a fully Al-sensitive background.

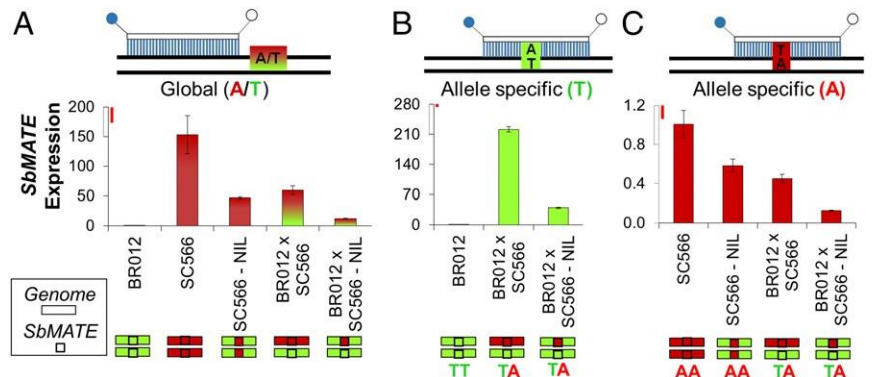
Trans-Acting Loci Influencing *SbMATE* Expression Are Present Within an *SbMATE* Expression/Al Tolerance QTL on Chromosome 9. Next, we undertook QTL mapping in a BR007 × SC283 recombinant inbred line (RIL) population. SC283 is highly tolerant to Al toxicity, whereas BR007 is highly Al-sensitive (Fig. 2A), and this RIL population was previously used to positionally clone *SbMATE* (4). The availability of a large population size for this highly contrasting cross (9) justified its choice for QTL mapping. Major QTL for both Al tolerance (Fig. 2B) and *SbMATE* expression (Fig. 2C)

were collocated with *SbMATE* on chromosome 3. This indicates that higher *SbMATE* expression and Al tolerance in SC283 (7) is achieved predominantly in *cis*. A comparatively smaller Al tolerance and *SbMATE* expression QTL (eQTL) was detected at ~51 to ~54 Mb [$2 < -\log_{10}(p) < 8$ for the eQTL] on chromosome 9 (zoomed in; Fig. 2B and C). Although other loci with similar *P* values are found elsewhere, the chromosome 9 QTL was chosen because of its clear joint effect on both Al tolerance and *SbMATE* expression. Multilocus mapping was also undertaken and revealed a possible interaction between the QTL on chromosomes 3 and 9 (*SI Appendix*, Table S1).

Genome-wide association mapping (GWAS) identified SNPs associated both with Al tolerance and *SbMATE* expression in the Al tolerance/eQTL region on chromosome 9 (Fig. 2D and E). In that region, SNP loci in linkage disequilibrium were detected across distances exceeding 2 Mb (*SI Appendix*, Fig. S1). Next, we identified SNPs associated with both Al tolerance and *SbMATE* expression, with $-\log(p)$ between 3 and 9. We overlapped the resulting physical interval with that of the QTL identified in the RIL population, and defined an extended region between ~45 and ~59 Mb (Fig. 2F) as a search region to identify candidate genes with regulatory signatures (<https://phytozome.jgi.doe.gov/pz/portal.html>, v1.4).

***SbMATE* Promoter.** The structure of the *SbMATE* promoter region that contains the MITE insertion is shown in *SI Appendix*, Fig. S2A and B. The MITE element (unit “b”) is flanked by 100-bp (unit “a”) and 20-bp (unit “c”) sequences. This MITE-containing a-b-c triplet (designated hereafter simply as “MITE repeats”) is followed by a single terminal (unrepeated) 100-bp “a” unit with either an 8-bp deletion (present in SC283 and Tx430) or a 12-bp deletion (present in BR012; *SI Appendix*, Fig. S2B). Henceforth, the 100-bp “a” sequence within the MITE repeats will be designated as the 100-bp repeat and the terminal, unrepeated units, as the 88- or 92-bp terminal. Natural, allelic variation at the *SbMATE* promoter arises from tandem variations in the number of identical a-b-c units, which are present either as a singlet or as repeated units in different sorghum lines, with Al-tolerant lines showing in general more repeats compared with Al-sensitive lines (4). For example, the parents of the RIL population, BR007 (Al-sensitive) and SC283 (tolerant), have three and five repeats, respectively. For trans-activation assays, we synthesized *SbMATE* promoters containing one MITE repeat (promoter from the Al-sensitive line Tx430, designated as Tx430p), four repeats (BR012p, from BR012, which is Al-sensitive), and five repeats (SC283p, from SC283, Al-tolerant), with the flanking, 1,749-bp and 2,010-bp sequences from the sorghum BAC where *SbMATE* resides (4). Promoter truncations were amplified from SC283.

Fig. 1. Global and allele-specific expression of *SbMATE* assessed with TaqMan probes. The global assay assesses the joint expression of *SbMATE* alleles, and allele-specific expression was based on an SNP (T, present in Al-sensitive BR012; A, present in Al-tolerant SC566) within *SbMATE*. Colored schematics indicate the genetic backgrounds (genome, rectangles) and the *SbMATE* alleles (squares): SC566 (red) and BR012 (green). (A) Global expression, allele-specific expression of the (B) Al-sensitive (T, green) allele and (C) Al-tolerant (A, red) allele, with the probes depicted on *Top*. The red/green gradient in A shows the joint expression of both the A and T alleles in stocks heterozygous for *SbMATE*. Global relative expression values are fold changes relative to the Al-sensitive line. Expression of the T and A alleles are fold changes relative to expression in the parents, BR012 and SC566, which are homozygous for the T and A alleles, respectively. The sorghum genotypes were grown with {27} μM Al^{3+} in nutrient solution at pH 4.0 for 5 d, and the root apex (1 cm) was collected for RNA isolation. Values are mean \pm SD; $n = 3$. Least significant difference (Fisher's LSD, $\alpha = 0.05$) bars (in red) are drawn to scale (*Top* of the y axis).



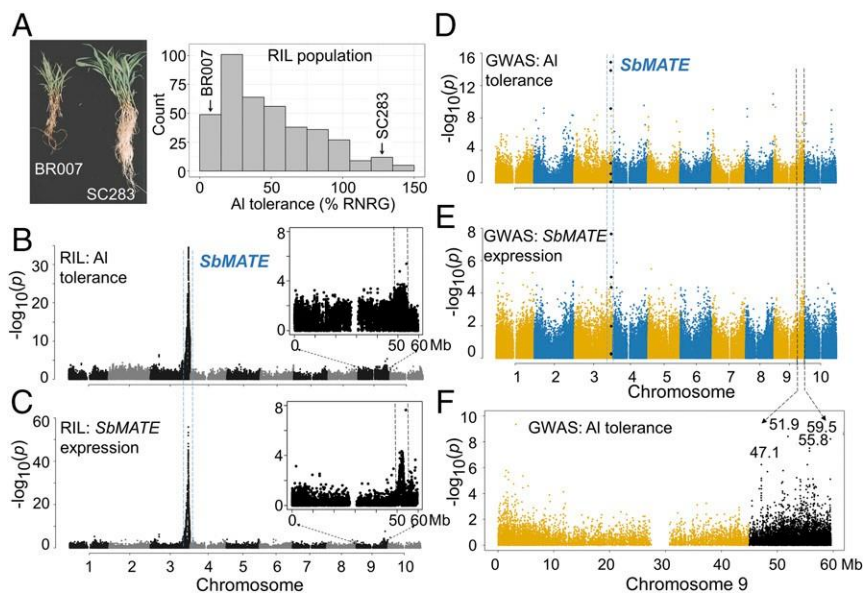


Fig. 2. *Trans-factor positional cloning based on expression GWAS.* (A) Contrasting phenotypes in BR007 (Al-sensitive) and SC283 (Al-tolerant) based on root damage and relative net root growth (%RNRG) assessed in nutrient solution with $\{27\} \mu\text{M Al}^{3+}$, in the context of the BR007 \times SC283 RILs. QTL mapping was carried out for (B) Al-tolerance and (C) *SbMATE* expression (eQTL) in the BR007 \times SC283 RILs. The chromosome 9 region, where a collocated Al-tolerance/eQTL was detected, is expanded. GWAS with (D) Al tolerance and (E) *SbMATE* expression. SNPs near or within *SbMATE* previously associated with Al tolerance (20) are depicted as black diamonds. Blue and black dashed lines indicate the *SbMATE* region (chr 3) and the region harboring significant SNPs on chromosome 9, respectively, which overlap in the RIL QTL map (B and C) and in the GWAS plots (D and E). (F) Details of the chromosome 9 region showing physical positions for SNPs with the strongest association signals with Al tolerance (black dots).

Transcription Factors on the Chromosome 9 QTL Transactivate the *SbMATE* Promoter in Yeast. A qualitative analysis based on the yeast one-hybrid assay indicated that, in the Al tolerance/eQTL region, a WRKY-like transcription factor (TF), *Sb09g023500* (*SbWRKY1*) at 53.14 Mb, and *Sb09g021530*, a gene encoding a zinc finger DHHC (zf-DHHC) domain-containing protein (*SbZNF1*) at 50.98 Mb (*SI Appendix*, Fig. S3 A and B), were both capable of *trans*-activating *SbMATE* promoter alleles harboring one, four, and five copies of the MITE repeats (*SI Appendix*, Fig. S2 C and D). In contrast, *SbNFY1*, a NFY-like TF (*Sb09g022810*) located within the same QTL, did not activate the *SbMATE* promoter, confirming the specificity of *SbWRKY1* and *SbZNF1* transcriptional activation of *SbMATE*.

Qualitative promoter deletion analysis (*SI Appendix*, Fig. S2 C and D) showed that a proximal SC283 promoter fragment, extending to position -2102 relative to the *SbMATE* start codon (-2102pSC283 , where “p” stands for promoter), was sufficient for *trans*-activation by both TFs, but *trans*-activation was lost when the 92-bp terminal was deleted (-2010pSC283).

SbZNF1 and *SbWRKY1* Bind Both to the 100-bp Repeat and to the 92-bp Terminal. Leaf protoplasts from transgenic *Arabidopsis* transformed with constructs containing the first 2,010 bp of the *SbMATE* promoter including either the 92-bp terminal (-2102p , from the SC283 promoter) or the MITE repeats in BR012 (-5299p ; Fig. 3 A and B) were transformed along with constructs encoding 35S-driven YFP::TF cDNA. The immunoprecipitated chromatin (ChIP) fragments obtained with anti-GFP antibody were analyzed by PCR and qPCR. For both *SbWRKY1*-expressing and *SbZNF1*-expressing protoplasts, qPCR with IP DNA showed that a fragment within the 92-bp terminal, amplified with primers F1/R4 and F1/R5, was significantly enriched over the input (control) DNA, confirming binding to the 92-bp terminal (Fig. 3A). In contrast, no enrichment was observed with primers annealing either to the 2,010-bp fragment (F2/R3), which lacks both the 100-bp repeat and the 92-bp terminal, or the actin gene (endogenous control).

ChIP was also undertaken with the -5299 promoter from BR012, which has four MITE repeats followed by the terminal 88-bp fragment (Fig. 3B). Amplification of IP DNA with primers F1/R1, which are specific to the 100-bp repeat, was significantly enriched over the input DNA (Fig. 3B). Collectively, the results in Fig. 3 A and B show that *SbWRKY1* and *SbZNF1* bind both to the 100-bp “a” unit within the MITE repeats and to the 92-bp terminal. The amplification profiles of IP DNA (*SI Appendix*, Fig. S4) with primers

flanking (F3 and F4) and within (R1) the 100-bp repeat (lanes 1 and 3 in -5299p) confirm such binding. Because the R1 primer in Fig. 3B does not anneal to the 88-bp terminal in -5299p , we cannot rule out that the TFs do not bind to that fragment because of its additional 4-bp deletion compared with the 92-bp terminal fragment in -2102p (see *SI Appendix*, Fig. S2B for “a” unit alignments).

The Number of MITE Repeats Correlates With Enhanced *SbWRKY1* and *SbZNF1* Transactivation Activity. Because transactivation assays in yeast are qualitative, we quantified transactivation activity in *Arabidopsis*. Protoplasts were isolated from *Arabidopsis* transformed with the truncated -2102 promoter from Al-tolerant SC283 containing the 92-bp terminal (Fig. 3C). In addition, promoter alleles containing one (Tx430p) and four (BR012p) copies of the MITE repeats followed by the 92-bp and 88-bp terminal (Fig. 3D), respectively, were tested. Reporter gene-specific activity was higher with protoplasts isolated from *Arabidopsis* cotransformed with the 92-bp terminal and with *SbWRKY1* and *SbZNF1* compared with the negative controls (Fig. 3C). In addition, elimination of the 92-bp terminal abolished reporter gene activity (*SI Appendix*, Fig. S5 A–H), which is consistent with the transactivation results in yeast.

Reporter gene activity was significantly higher with promoter alleles containing four (BR012p) compared with one (Tx430p) copies of the MITE repeats (Fig. 3D). For both promoter constructs, a cotransformation assay indicated a synergistic mode of action, with a greater effect on *SbMATE* promoter activity when both TFs are present (*SbWRKY1* + *SbZNF1*) compared with their individual effects (Fig. 3D). These two promoter alleles vary both for the number of MITE repeats and for the presence of an additional 4-bp deletion specifically in the 88-bp unique terminal of the four-repeat promoter. However, we established previously that both TFs bind to the 100-bp “a” repeat and to the unrepeated 92-bp “a” terminal (Fig. 3 A and B). In view of that, even if both TFs did not bind to the terminal “a” 88-bp fragment of the four MITE-repeat promoter, this promoter would still harbor four binding sites for the TFs within its MITE region, in contrast to only two binding sites in the one MITE-repeat promoter (within its single 100-bp sequence and in the 92-bp terminal). It is thus unlikely that the additional 4-bp deletion was the cause of the higher transactivation in the 4 MITE-repeat (BR012p) promoter compared with the 1 MITE-repeat promoter (Tx430p). These results strongly suggest that increased binding site abundance in promoters where the number of MITE repeats has expanded leads to enhanced TF recruitment.

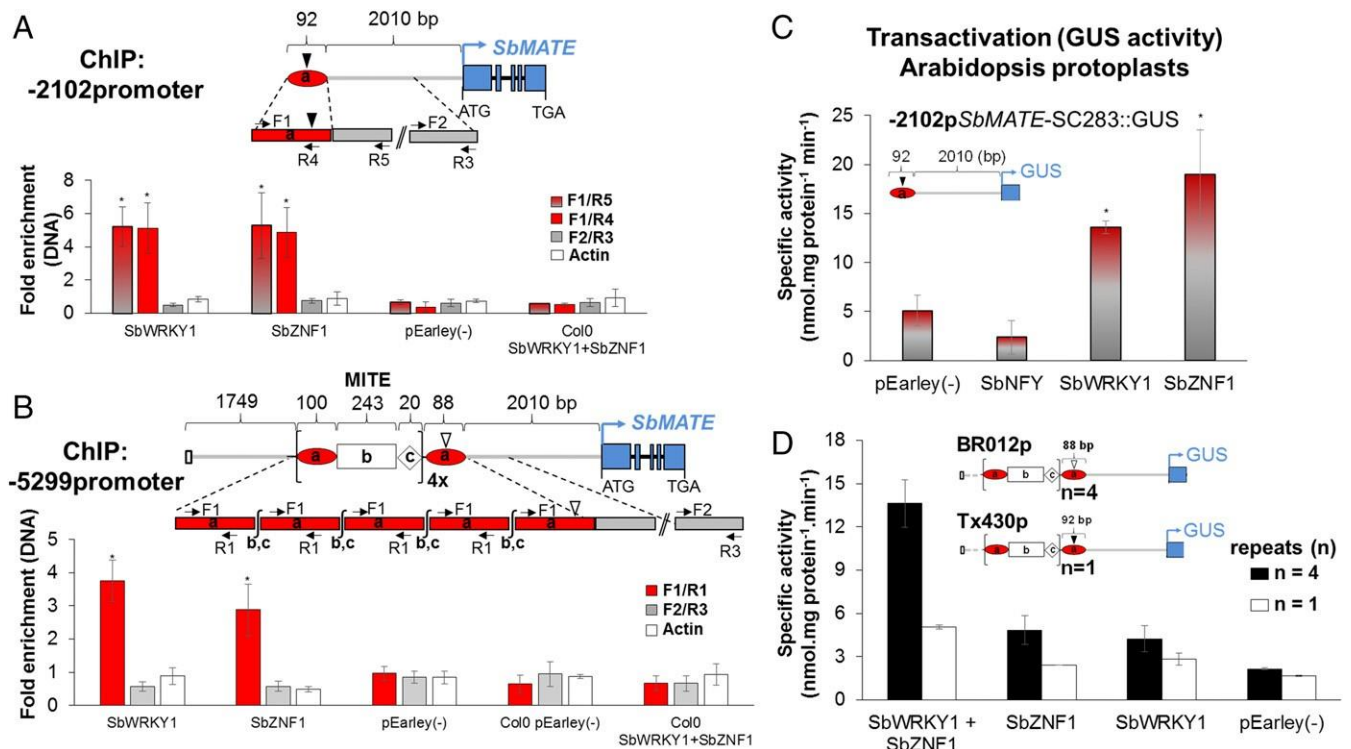


Fig. 3. *SbWRKY1* and *SbZNF1* bind to and transactivate the *SbMATE* promoter. For both (A and B) ChIP-qPCR and (C and D) transactivation assays, protoplasts from *Arabidopsis* stably transformed with promoter constructs were transfected with pEarleyGate104 containing the indicated TFs or with the empty vector (minus sign). Schematics of the *SbMATE* promoter constructs are shown: the general structure MITE region is depicted in B: the 243-bp MITE element (unit “b”) is flanked by 100-bp (unit “a”, 100-bp repeat) and 20-bp (unit “c”) sequences, and the number of identical a-b-c triplets varies in different promoter alleles. The MITE-containing a-b-c triplets (MITE repeats) is terminated by a single 100-bp “a” unit with either an 8-bp deletion (depicted by an inverted black triangle, present in SC283 and Tx430) or a 12-bp deletion (inverted white triangle, present in BR012), which results in the 88- or 92-bp terminal fragments. ChIP-qPCR with (A) a truncated *SbMATE* promoter from the SC283 line extending to position -2,102 bp [-2102promoter(p)] and (B) from BR012 extending to position -5,299 bp [-5299promoter(p)] and containing four copies of the MITE repeats (4x), followed by the 88-bp terminal. Arrows represent primer (SI Appendix, Table S2) positions. Data were normalized to the input (control) for each sample and are expressed as the fold-enrichment vs. preimmune IgG serum controls. Error bars indicate SEM ($n = 3$), and asterisks indicate significant differences ($P < 0.05$). (C and D) Transactivation in *Arabidopsis* protoplasts. Protoplasts stably transformed with (C) -2102pSC283::GUS and transfected with pEarleyGate104 with or without [pEarley(-)] the sorghum TFs. Results are the mean \pm SD of three independent experiments. * $P < 0.05$; Scott-Knott test. (D) Transactivation with the promoter region of Tx430 (-4214pTx430::GUS) and BR012 (-5299pBR012::GUS), with one and four MITE repeats (n), respectively. The error bars represent the limits of the non-parametric bootstrap confidence interval of 95% with $n = 4$.

SbZNF1 and *SbWRKY1* Alleles from Al-Tolerant and Al-Sensitive Lines Are Differentially Regulated by Al³⁺. For clarity, alleles are designated in the text with the gene names (TF is used when referring to both *SbWRKY1* and *SbZNF1*) subscribed with numbers indicating the allele donors. We looked at the expression profiles of *SbWRKY1* and *SbZNF1* alleles derived from unrelated Al-tolerant (SC566 and SC283) and Al-sensitive (BR007 and BR012) lines (10) (Fig. 4A). Genetic backgrounds are depicted by the colored rectangles beneath Fig. 4A. The SC566- and SC283-NILs carry respective *SbMATE* allele (*SbMATE*_{566/283}, depicted by squares) in the BR012 genetic background, and as such, their TF alleles (green ovals) are the same as the ones in BR012 (*TF*₀₁₂).

In the presence of Al³⁺, both *SbWRKY1* and *SbZNF1* were more highly expressed in Al-tolerant (SC283 and SC566) compared with Al-sensitive (BR012 and BR007; Fig. 4A) lines. In NILs in which tolerant *SbMATE* alleles (*SbMATE*_{283/566}) were introgressed into the Al-sensitive BR012 background (SC283- and SC566-NILs), TF expression was reduced compared with their respective Al-tolerant donors (SC283 and SC566). These responses are similar to their transcriptional target *SbMATE*, which also showed reduced expression in the SC566- and SC283-NILs compared with the Al-tolerant parents (SI Appendix, Fig. S6, 5 d, and ref. 7). Strikingly, the *SbWRKY1* allele derived from the Al-tolerant line SC566, *SbWRKY1*₅₆₆, was markedly up-regulated by

Al³⁺ (*SbWRKY1*₂₈₃ also shows a consistent tendency for Al³⁺ up-regulation, but slighter). In contrast, the Al-sensitive alleles in BR007 (*SbWRKY1*₀₀₇), BR012, and NILs (*SbWRKY1*₀₁₂) were strongly down-regulated by Al³⁺ (Fig. 4A). *SbWRKY1* and *SbZNF1* exhibited different transcriptional responses to Al³⁺ in different Al-tolerant lines, as Al-induced *SbWRKY1* expression was greater in SC566 compared with SC283, whereas *SbZNF1* Al³⁺ up-regulation and expression was higher in SC283.

In Sorghum, *SbWRKY1* and *SbZNF1* Alleles Derived from SC283 (Al-Tolerant) Increase *SbMATE* Expression. A genetic analysis in the BR007 \times SC283 RIL population was conducted using *SbWRKY1* (W) and *SbZNF1* (Z) gene-specific markers, which were designed based on polymorphisms that differentiate the *TF*₂₈₃ and *TF*₀₀₇ alleles (Fig. 4B). This was done to select RILs with different combinations between parental alleles of *SbZNF1* and *SbWRKY1* (i.e., TF haplotypes). For this analysis, we compared *SbMATE* expression and Al tolerance of RILs selected to contain both TF alleles from the Al-tolerant parent (*TF*₂₈₃, Z⁺/W⁺), from the Al-sensitive parent (*TF*₀₀₇, Z⁻/W⁻), and one TF allele from each parent (Z⁺/W⁻, Z⁻/W⁺).

A linear regression model fit to all haplotype classes indicated that both TFs enhanced *SbMATE* expression (Fig. 4C and SI Appendix, Fig. S7). *SbMATE* expression for the double homozygous

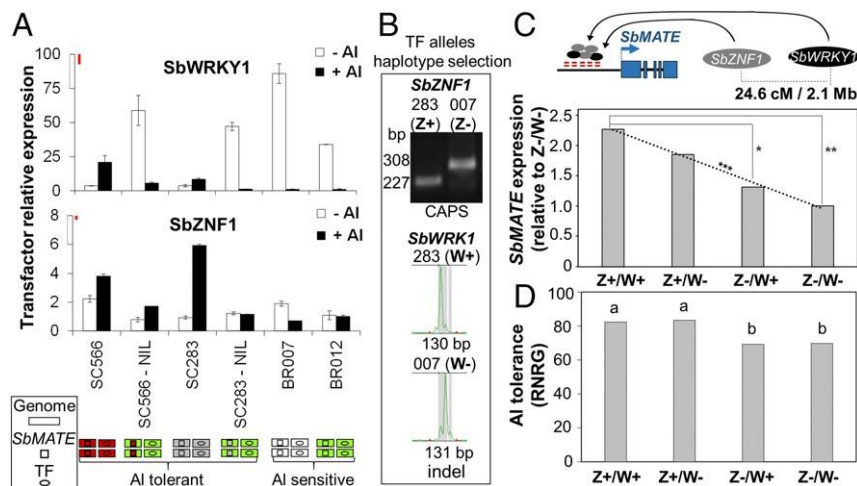


Fig. 4. Transcription factor expression profile and effect on *SbMATE* expression and Al tolerance. (A) *SbWRKY1* and *SbZNF1* expression in Al-tolerant (SC283 and SC566) and Al-sensitive (BR007 and BR012) lines, and in the SC566-NIL and SC283-NIL (SC566/SC283 *SbMATE* in the BR012 background). Colored schematics indicate the genetic backgrounds (genome, rectangles), the *SbMATE* alleles (squares), and TF alleles (ovals), along with the Al tolerance phenotype (7). Plants were grown on $\pm 27 \mu\text{M}$ Al^{3+} for 5 d in nutrient solution at pH 4.0 (brackets denote free Al^{3+} activity estimated with GEOCHEM; see *SI Appendix*, Supplementary Methods), and the root apices (1 cm) were collected. Values are mean \pm SD ($n = 2$). Least significant difference (Fisher's LSD, $\alpha = 0.10$) bars (in red) are drawn to scale (*Top* of the y axis). This experiment was repeated with similar results (*SI Appendix*, Fig. S6, 5 d; $n = 3$). (B) *SbWRKY1* and *SbZNF1* polymorphisms in the RIL parents, BR007 (007, Al-sensitive) and SC283 (283, Al-tolerant), which were used to select RILs with all combinations of TF alleles (i.e., TF haplotypes). (C) *SbWRKY1* and *SbZNF1* effect on *SbMATE* expression estimated based on RILs homozygous for the SC283 (Al-tolerant) alleles at both TF loci (Z^+/W^+), for the BR007 (Al-sensitive) allele (Z^-/W^-), or showing alternate TF alleles (Z^+/W^- and Z^-/W^+). Significant differences based on 5% (***) and 12% (*) confidence intervals. A linear regression model fit to haplotype *SbMATE* expression was highly significant (***) $\alpha = 0.01$). Physical (Mb) and genetic (cM) distances between TFs are depicted at the *Top*. (D) Effect of *SbWRKY1* and *SbZNF1* on Al tolerance as measured by relative net root growth (%NRNG). Different letters indicate statistical differences (Fisher's least significant difference, $\alpha = 0.08$).

haplotypic class containing the Al-tolerant SC283 allele at both loci (Z^+/W^+) produced the largest increase in *SbMATE* expression compared with the other haplotype classes, and was more than 2.3-fold higher than the class containing Al-sensitive alleles from BR007 at both loci (Z^-/W^- ; Fig. 4C).

SbZNF1 exerted a stronger effect on both *SbMATE* expression (Fig. 4C) and Al tolerance (Fig. 4D) compared with *SbWRKY1*, the individual effect of which on Al tolerance was below the statistical power of our haplotype-based approach. This result is likely population-specific, resulting from stronger Al^{3+} up-regulation of *SbZNF1* expression compared with *SbWRKY1* specifically in the Al-tolerant parent of the RIL population, SC283 (Fig. 4A).

Time-Dependent Expression in Root Apices Exposed to Al^{3+} for *SbWRKY1* and *SbZNF1* Is Similar to *SbMATE*. Al-induced expression of both *SbWRKY1* and *SbZNF1* was higher in root apices of Al-tolerant lines compared with the rest of the root system and shoots (*SI Appendix*, Fig. S8). This response favoring preferential expression in root apices was larger for *SbWRKY1* in SC566 compared with *SbZNF1* in SC283, which are the genotypes that display the highest expression of each TF gene under Al^{3+} (Fig. 4A). A time-course analysis indicated a general trend for time-dependent increase in TF expression in Al-tolerant lines between 1 and 5 d of Al exposure, which was higher for *SbWRKY1* (4.2–4.8-fold) compared with *SbZNF1* (1.1–1.6-fold). In general, preferential, time-dependent expression in root apices exposed to Al^{3+} for *SbZNF1* and *SbWRKY1* parallels the *SbMATE* expression measured under the same period in Al. For the Al-sensitive lines, BR007 and BR012, *SbWRKY1* and *SbZNF1* expression in the presence of Al^{3+} decreased over the same 1-, 3-, and 5-d periods (*SI Appendix*, Fig. S6).

Discussion

We discovered that *SbMATE* expression is influenced by a *cis*-acting tandemly repeated sequence flanking a MITE insertion upstream of *SbMATE*, which provides sites in which *SbWRKY1* and *SbZNF1* bind and transcriptionally regulate *SbMATE*.

Possible binding motifs in the binding fragment are the recognition core for Dof1/MNB1a zf-TFs (11,12) and a motif similar to the WT-box, where a WRKY TF has been shown to bind (13) (see *SI Appendix*, Table S3 for *cis* elements identified in silico).

Our results indicate that *SbMATE* and *SbWRKY1* are coregulated ($r = 0.3$; $P = 0.08$; *SI Appendix*, Fig. S6). This suggests that *SbWRKY1* functionally evolved to regulate *SbMATE* expression in response to Al^{3+} , which is consistent with the active and adaptable nature of Group III C2H-type zfs (14). *SbZNF1* is a DHHC-like S-acyl transferase zf, and such proteins have been implicated in abiotic stress tolerance (15). *SbZNF1* is preferentially expressed in roots of Al-tolerant lines, but its expression is localized to the root tip to a lesser extent than *SbWRKY1*. This pattern may reflect the more general physiological role of DHHC proteins, stemming from the DHHC cognate function in increasing protein hydrophobicity (16). Our quantitative analysis of transactivation in *Arabidopsis* protoplasts positively associated *SbWRKY1* and *SbZNF1* transactivation activity and the number of MITE repeats in the *SbMATE* promoter, suggesting a dosage dependency. Hence, we propose that the singular (17), tandemly repeated structure of the MITE repeats has led to differential TF recruitment (*SI Appendix*, Fig. S9 A and B), resulting in the previously observed positive correlation between the size of the MITE insertion region and the Al tolerance phenotype (4).

Synergistic transactivation in *Arabidopsis* protoplasts, in conjunction with our haplotype analysis of *SbWRKY1* and *SbZNF1* in RILs derived from parents harboring different TF alleles, suggest that, in sorghum, these TFs cooperate to increase *SbMATE* expression. The cDNA sequences of *SbWRKY1* and *SbZNF1* alleles in Al-tolerant and Al-sensitive lines were found to be identical. Therefore, differential, time-dependent regulation by Al^{3+} of TF alleles appears to be a critical step in the *cis/trans* interactions that control *SbMATE* expression. Although the SC283 (Al-tolerant) allele of both TFs is up-regulated by Al^{3+} , the alternative, BR007 (Al-sensitive allele), is down-regulated (Fig. 4A), which helps to explain why RILs fixed for the SC283 alleles of both *SbWRKY1* and *SbZNF1* show a 2.3-fold increase in *SbMATE* expression compared

with RILs fixed for the BR007 alleles (Fig. 4C). The estimated TF effect on Al tolerance (18% increase; Fig. 4D: Z^+/W^+ vs. Z^-/W^-) is equal to the decrease in Al tolerance when the SC283 allele of *SbMATE* was introgressed into the background of the Al-sensitive line, BR012 (~18% in SC283 vs. SC283-NIL; figure 2 in ref. 7), which we show here has low-expressing alleles for both *SbWRKY1* and *SbZNF1* (Fig. 4A). This suggests that allelic variation at the TF loci is responsible for our previously observed genetic background effects, which lead to reduced expression of Al-tolerant alleles of *SbMATE* when introgressed into Al-sensitive backgrounds (7).

Our *cis/trans* interaction model (SI Appendix, Fig. S9) depicts a possible compensatory mode of action for *cis* and *trans* effects in highly Al-tolerant lines. Accordingly, the loss of one MITE repeat in SC566 compared with SC283 would be expected to reduce TF occupancy and, hence, reduce *SbMATE* expression in SC566. However, the loss of one binding unit appears to be compensated for by higher Al^{3+} up-regulation of *SbWRKY1* expression in SC566, resulting in higher *SbMATE* expression in the presence of Al^{3+} . Conversely, in the absence of Al^{3+} , high expression of *SbWRKY1* and *SbZNF1* in Al-sensitive lines (BR007 and BR012, with three and four MITE repeats, respectively; SI Appendix, Fig. S9) does not lead to enhanced *SbMATE* expression (Fig. 4A and SI Appendix, Fig. S6). This suggests the occurrence of independent *cis*-acting repressor components acting upstream of the MITE repeats in their role of providing TF binding sites. Although these components likely control extreme Al tolerance and sensitive phenotypes, *SbWRKY1* and *SbZNF1* appear to act to regulate *SbMATE* in Al-tolerant lines.

Compensatory *cis/trans* effects (2, 18), which appear to be a rather widespread mechanism that cells use to stabilize gene expression (19), are implicated in coevolution between *cis* and *trans* mutations. Although *cis* variants may provide a more stable control of gene expression under stress, *trans* regulation is important for environmental responses (2). In the absence of Al^{3+} , both *SbWRKY1* and *SbZNF1* are down-regulated in Al-tolerant genotypes (SC283 and SC566), reducing *SbMATE* expression precisely when root citrate release, which can be costly to the plant, is not needed because of the lack of Al toxicity. Therefore, the interplay between *cis*-acting elements and TFs that are responsive to Al^{3+} stress may be advantageous, as a result of a balancing effect on *SbMATE* expression, which would otherwise be more inflexibly controlled in *cis*, resulting in genetic load related to the loss of carbon during unnecessary root citrate release.

In light of the molecular nature of *cis* and *trans* variants that modulate *SbMATE* expression, we can now both predict and circumvent genetic background effects that reduce *SbMATE* expression to increase grain yield production on acidic, Al toxic soils across the world.

Materials and Methods

Genetic Stocks. Development of NILs, RILs, and hybrid stocks, and the association panel used for GWAS (20), are described in SI Appendix, Supplementary Methods.

Al Tolerance in Hydroponics. Al tolerance was assessed based on root growth inhibition, relative net root growth (RNRG), in nutrient solution with and without $\{27\} \mu M Al^{3+}$ at pH 4.0 (20) (brackets denote free Al^{3+} activity estimated with GEOCHEM; see SI Appendix, Supplementary Methods).

Gene Expression via Quantitative RT-PCR. Sorghum plants were grown in nutrient solution $\pm\{27\} \mu M Al^{3+}$ for 1 and/or 3 and 5 d, depending on the experiment. Gene expression was assessed either with the TaqMan Gene Expression or SYBR Green assay (Applied Biosystems). Allele-specific expression was assessed (TaqMan) based on an A/T SNP in the first exon of *SbMATE* (7), with the A allele present in SC566 and the T allele present in all other lines. See SI Appendix, Supplementary Methods.

QTL Mapping in a RIL Population. Al tolerance and *SbMATE* expression data were obtained in nutrient solution with $\{27\} \mu M Al^{3+}$ at pH 4.0 for 5 d, and QTL mapping with SNP markers was undertaken with TASSEL (GLM) and by multiple regression.

Genomewide Association Mapping. Genomewide association mapping was undertaken based on a mixed linear model (Q + K) with TASSEL. SNP markers were tested for associations with Al tolerance [RNRG; SI Appendix, Table S4 (20)] and *SbMATE* expression ($\Delta\Delta Ct$) at 5 d of Al exposure.

Transactivation Assays. Full-length promoter fragments and *trans*-factor cDNA (v1.4 of the sorghum genome) sequences were commercially synthesized, and transactivation assays were conducted as described in SI Appendix, Supplementary Methods. The experiments were repeated four times with similar results.

Transcription Factor Effects on *SbMATE* Expression and Al Tolerance via Haplotype Analysis in an RIL Population. *SbWRKY1* and *SbZNF1* genotyping was based on an indel and a SNP polymorphism, respectively, as described in the SI Appendix, Supplementary Methods.

Chromatin Immunoprecipitation Assay. Leaf protoplasts were isolated from transgenic *Arabidopsis thaliana* plants transformed with different *SbMATE* promoter fragments and then transformed with the 35S::YFP::*SbWRKY1* and 35S::YFP::*SbZNF1* vectors. See SI Appendix, Supplementary Methods.

ACKNOWLEDGMENTS. We thank Veridiana Cano for assistance with allele-specific expression assays and William Lucas (University of California, Davis) for critically reading the manuscript. We acknowledge grants from the CGIAR Generation Challenge Program, the Embrapa Macroprogram, the Fundação de Amparo a Pesquisa do Estado de Minas Gerais, and the National Council for Scientific and Technological Development.

- Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet* 40:346–350.
- Cubillos FA, et al. (2014) Extensive *cis*-regulatory variation robust to environmental perturbation in *Arabidopsis*. *Plant Cell* 26:4298–4310.
- von Uexküll HR, Mutert E (1995) Global extent, development and economic impact of acid soils. *Plant Soil* 171:1–15.
- Magalhaes JV, et al. (2007) A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat Genet* 39:1156–1161.
- Carvalho G, Jr, et al. (2016) Back to acid soil fields: The citrate transporter *SbMATE* is a major asset for sustainable grain yield for sorghum cultivated on acid soils. *G3 Genes Genomes Genet* 6:475–484.
- Leiser WL, et al. (2014) Two in one sweep: Aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African sorghum. *BMC Plant Biol* 14:206.
- Melo JO, et al. (2013) Incomplete transfer of accessory loci influencing *SbMATE* expression underlies genetic background effects for aluminum tolerance in sorghum. *Plant J* 73:276–288.
- Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. *Curr Opin Genet Dev* 5:814–821.
- Magalhaes JV, et al. (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the poaceae. *Genetics* 167:1905–1914.
- Caniato FF, et al. (2007) Genetic diversity for aluminum tolerance in sorghum. *Theor Appl Genet* 114:863–876.
- Yanagisawa S (1995) A novel DNA-binding domain that may form a single zinc finger motif. *Nucleic Acids Res* 23:3403–3410.
- Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *Plant J* 17:209–214.
- Machens F, Becker M, Umrath F, Hehl R (2014) Identification of a novel type of WRKY transcription factor binding site in elicitor-responsive *cis*-sequences from *Arabidopsis thaliana*. *Plant Mol Biol* 84:371–385.
- Huang Y, et al. (2016) Members of WRKY group III transcription factors are important in TYLCV defense signaling pathway in tomato (*Solanum lycopersicum*). *BMC Genomics* 17:788.
- Zhou L-Z, et al. (2013) Protein S-ACYL Transferase10 is critical for development and salt tolerance in *Arabidopsis*. *Plant Cell* 25:1093–1107.
- Chamberlain LH, Shipston MJ (2015) The physiology of protein S-acylation. *Physiol Rev* 95:341–376.
- Wang S, Zhang L, Meyer E, Matz MV (2010) Characterization of a group of MITEs with unusual features from two coral genomes. *PLoS One* 5:e10700.
- Kuo D, et al. (2010) Coevolution within a transcriptional network by compensatory *trans* and *cis* mutations. *Genome Res* 20:1672–1678.
- Goncalves A, et al. (2012) Extensive compensatory *cis-trans* regulation in the evolution of mouse gene expression. *Genome Res* 22:2376–2384.
- Caniato FF, et al. (2014) Association mapping provides insights into the origin and the fine structure of the sorghum aluminum tolerance locus, *Alt_{ss}*. *PLoS One* 9:e87438.

SUPPLEMENTARY FIGURES

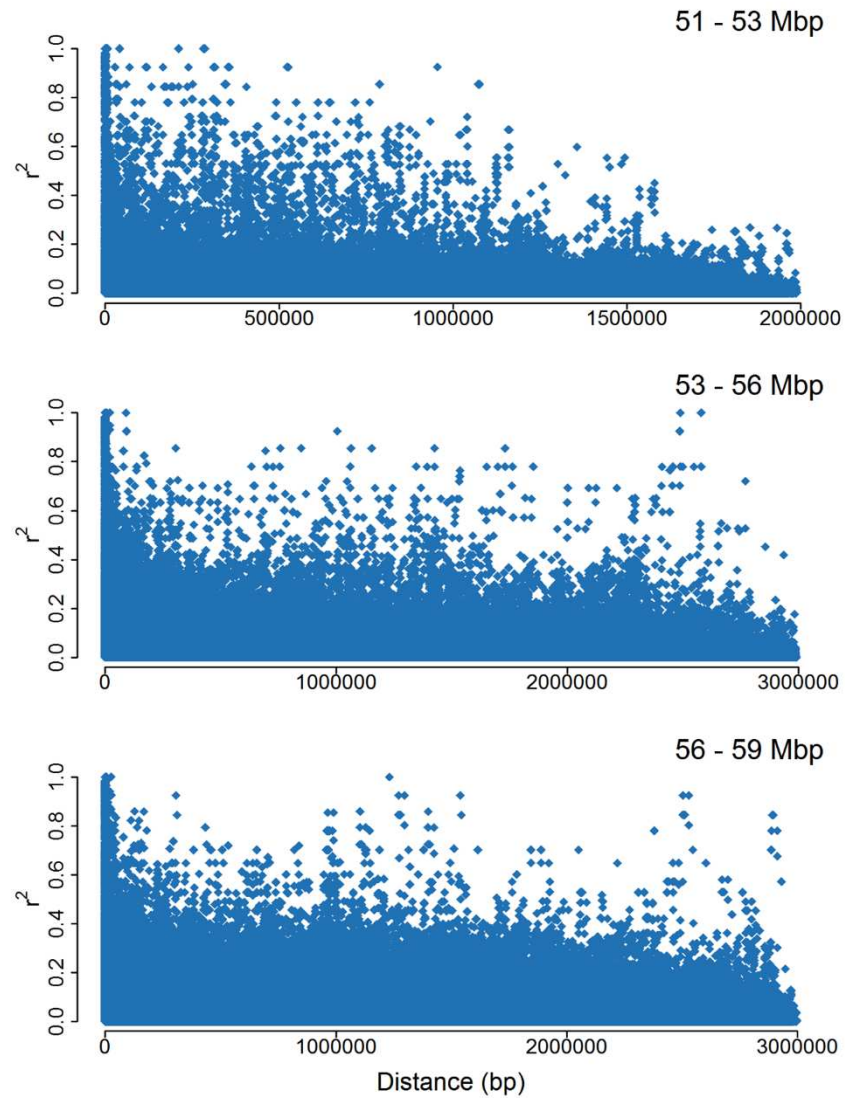


Fig. S1. Linkage disequilibrium (LD) at the end region of chromosome 9. The physical region where SNPs associated with AI tolerance and *SbMATE* expression were found (Fig. 2F) was assessed for LD. The squared allele-frequency correlations (21) (r^2) is shown as a function of physical distance between pairs of loci.

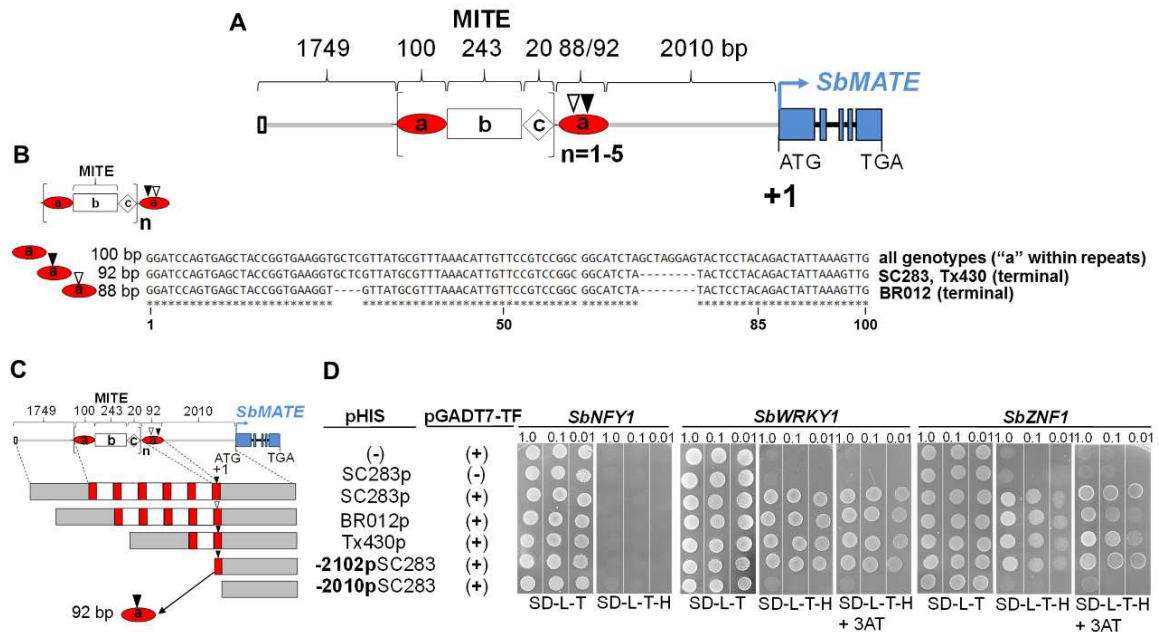


Fig. S2. *SbWRKY1* and *SbZNF1* activate *SbMATE* promoter alleles in yeast.

(A) Structure of the *SbMATE* promoter region. The 243-bp MITE element (unit "b") is flanked by 100-bp (unit "a", 100-bp repeat) and 20-bp (unit "c") sequences. This MITE-containing a-b-c triplet (designated MITE repeats) is followed by a single copy of the 100-bp "a" unit with either an 8-bp deletion (depicted by an inverted black triangle, present in SC283 and Tx430) or a 12-bp deletion (inverted white triangle, present in BR012), which are the terminal 88- or 92-bp fragments depicted in (B). The number of identical a-b-c units (n), varies from 1 to 5 in different sorghum lines. (C) *SbMATE* promoter alleles (SC283p, BR012p and Tx430p) with 5, 4 and 1 repeats, respectively, and SC283 truncated constructs containing (-2102pSC283) or lacking (-2010pSC283) the 92-bp terminal were cloned into pHIS2.1. (D) Transactivation assays in yeast. The vector pGADT7 contained (+) either the nuclear transcription factor Y (*SbNFY1*, as a negative TF control), *SbWRKY1* or *SbZNF1* TFs (pGADT7-TF). Strains carrying empty plasmids (pGADT7(-) and pHIS(-)) are negative controls. Growth of co-transformed yeast in synthetic drop-out medium lacking leucine and tryptophan (SD-L-T) confirm the presence of both constructs. Transactivation (Y1H) was monitored on SD-L-T-H (H, histidine) with 10mM 3-AT. Transformant dilutions are shown at the top.

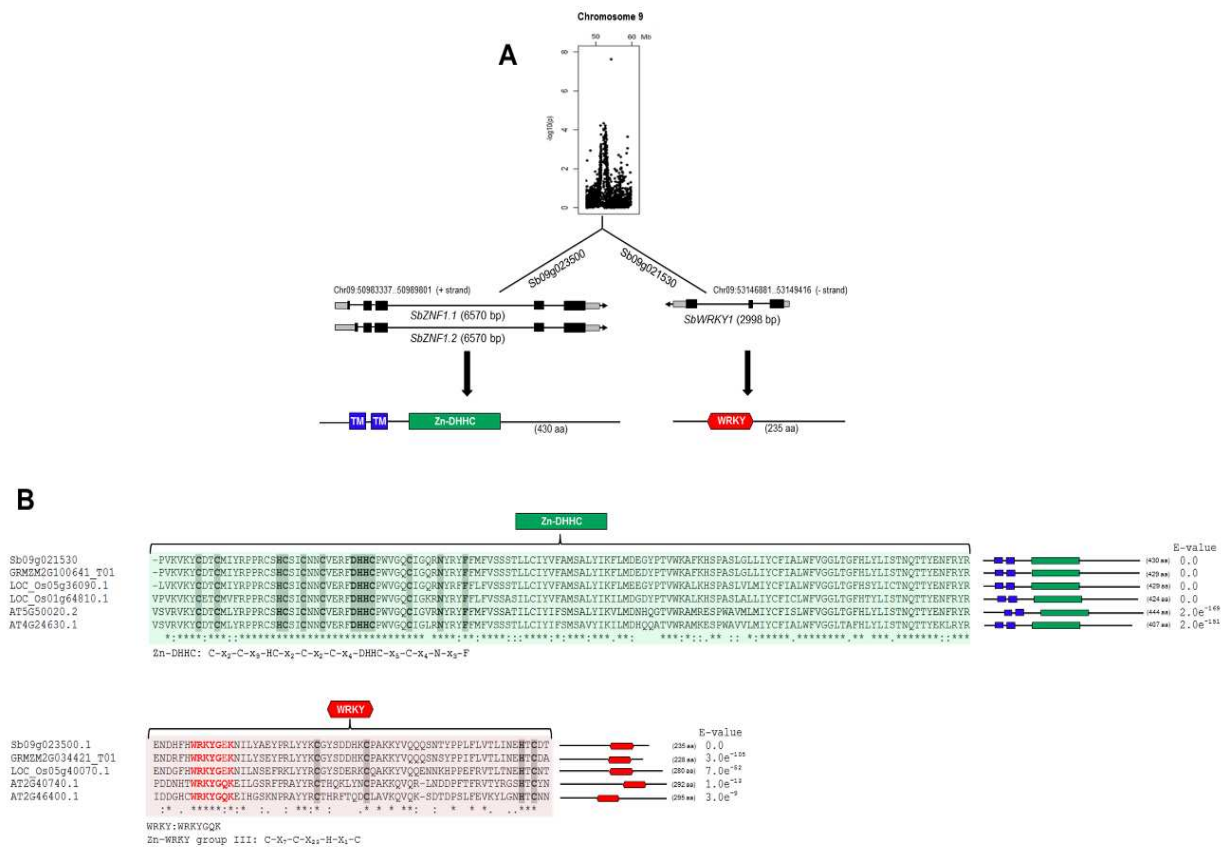


Fig. S3. *In silico* analysis of *SbWRKY1* and *SbZNF1*. (A) Physical positions of *SbZNF1* (*Sb09g023500*) and *SbWRKY1* (*Sb09g021530*) in the context of the *SbMATE* expression QTL on chromosome 9 (Fig. 2B) and gene models based on *Sorghum bicolor* v3.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>). This version predicts an isoform of *SbZNF1* (*SbZNF1.2*) in addition to *SbZNF1.1*, which is common to both v1.4 and v3.1 of the sorghum genome. Protein domains were predicted with Pfam (<http://pfam.xfam.org/>). The positions of the transmembrane (TM), Zn-DHHC and WRKY domains are shown. (B) Multiple alignment of the Zn-DHHC and WRKY domains was undertaken with Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and domain architectures of related proteins in maize (GRMZM), rice (LOC_Os) and Arabidopsis (AT) are shown. Expected values (*e-values*) were obtained based on a sequence similarity analysis using as query the complete protein sequences of *SbWRKY1* and *SbZNF1* in Phytozome. Alignment of the WRKY domain indicated that *SbWRKY1*, *AtWRKY46* (AT2G46400) and the remaining proteins possess a single C2H-type zinc-finger motif, typical of WRKY Group III (22). The WRKY (22) and the Zn-DHHC consensus (Pfam, PFM01529) are shown. Identical (*) amino acid residues and conservation between groups of strongly (:) or weakly (.) similar properties are shown in the ClustalO output.

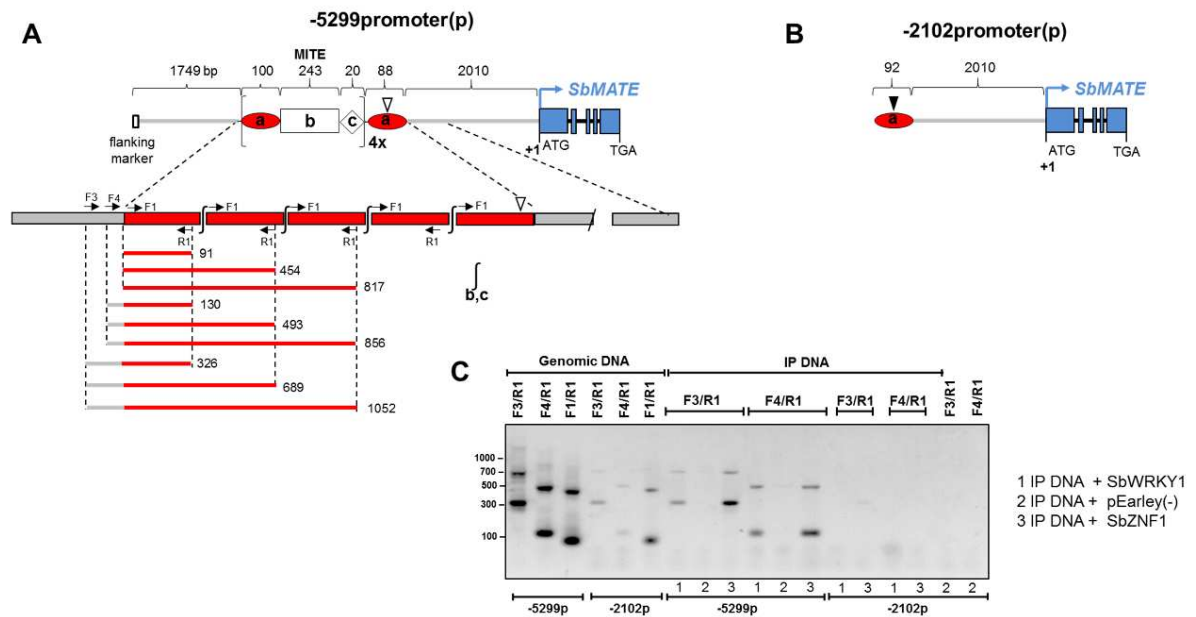


Fig. S4. Chromatin Immunoprecipitation (ChIP) assays show that SbWRKY1 and SbZNF1 bind to the 100-bp a repeat (within the MITE repeats). Schematic representation of the *SbMATE* promoters used in the ChIP assays: (A) promoter fragment from the BR012 line extending to position -5,299-bp [-5299promoter(p)] relative to the *SbMATE* start codon, and (B) SC283 promoter fragment extending to position -2,102-bp relative to the *SbMATE* start codon [-2102promoter(p)]. White and black triangles on the terminal “a” unit in BR012 and SC283 depict a 12-bp deletion in BR012 and 8-bp deletion in SC283, compared to the full 100-bp a unit within the MITE. Arrows represent primer positions whose sequences are in Table S2. The forward primers, F3 and F4, anneal to the unique 1749-bp fragment upstream of the 100-bp repeat whereas the R1 primer anneals within this 100-bp repeat. The expected sizes of each PCR-amplified fragments are shown as numbers in front of thick red and gray lines. (C) PCR analysis of immunoprecipitated (IP) DNA from *Arabidopsis* protoplasts transfected with vector pEarleyGate104 containing SbWRKY1 or SbZNF1 inserts, or the empty vector (minus sign). Genomic DNA from *A. thaliana* plants transformed with the promoter constructs was used as positive control.

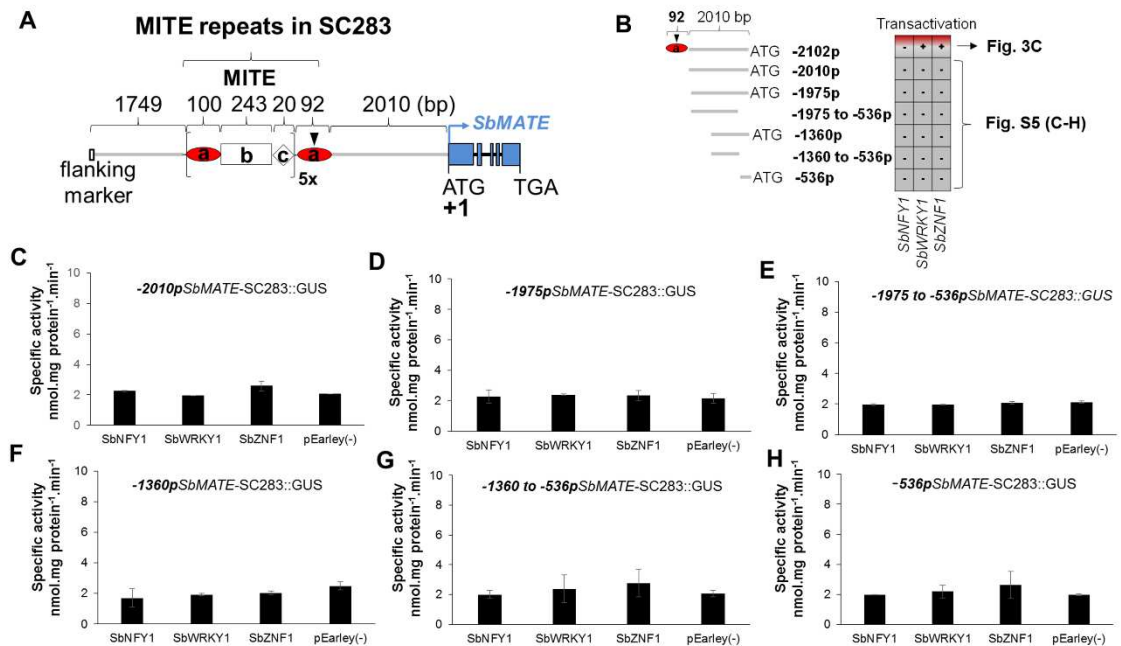


Fig. S5. Transactivation analysis of different truncated fragments of the *SbMATE* promoter by SbWRKY1 and SbZNF1. (A) Structure of the MITE repeats as described in Fig. 3. Protoplasts obtained from *Arabidopsis* were stably transformed with the GUS-fused promoter constructs and then electroporated with plasmids containing SbNFY1, SbWRKY1, SbZNF1, or the empty pEarleyGate104 (pEarley(-)) vector under the control of the S35 promoter. (B) Schematic representation of the transactivation results with a construct containing the terminal 92-bp fragment (data in Fig. 3C) or in truncated constructs excluding this terminal fragment (C – H). After 36 hours, β -glucuronidase activity was determined from protein extracts isolated from the transfected cells. The activities are shown as the means \pm S.D. from three independent experiments, each of which was performed with triplicate samples. The transactivation activity between the empty vector and the TFs for each promoter:GUS construct is not statistically different. The *SbMATE* promoter of SC283 was used to construct promoter truncations.

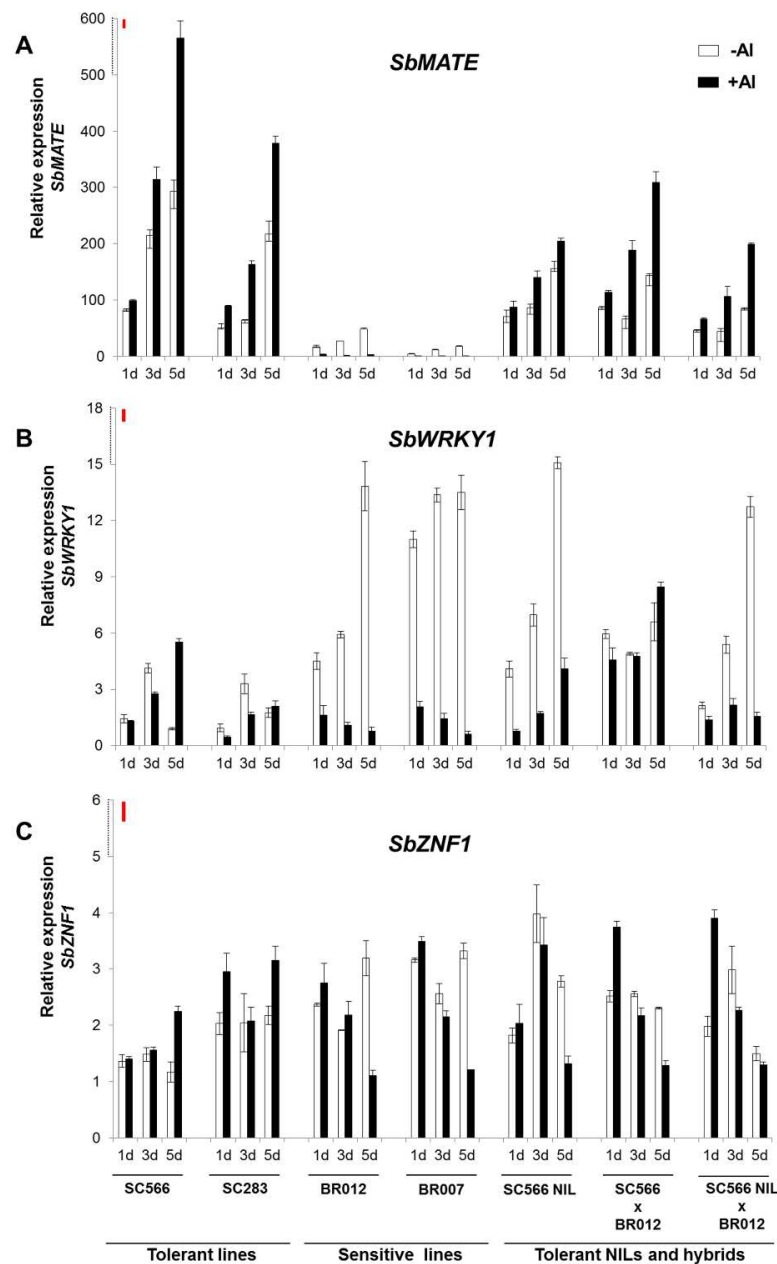


Fig. S6. Temporal expression of *SbMATE* and interacting transcription factors. (A) *SbMATE*, (B) *SbWRKY1* and (C) *SbZNF1* expression in SC566 (Al-tolerant), BR012 (Al-sensitive), SC566 - NIL (BR012 genetic background) and the hybrids BR012 x SC566-NIL and BR012 x SC566. All these genotypes possess 4 MITE repeats. The parents of our RIL population, SC283 (Al-tolerant) and BR007 (Al-sensitive), with 5 and 3 MITE copies, respectively, are included. The sorghum genotypes were grown without (-Al) or with {27} μM Al^{3+} (+Al) in nutrient solution at pH 4.0 for 1, 3 and 5 days and root apices (1 cm) were collected for RNA isolation. Error bars depicting standard deviation ($n=3$) are shown and least significant difference (Fisher's LSD, $\alpha = 0.05$) bars (in red) are drawn to scale for each gene (top of the Y-axis). A Pearson correlation analysis was undertaken with gene expression excluding SC283 and BR007, so that only genotypes with 4 MITE copies were used. *SbWRKY1* expression was significantly correlated with *SbMATE* expression ($r = 0.32$, $p = 0.08$).

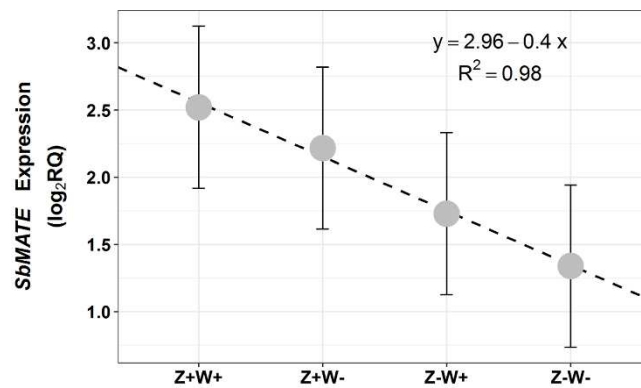


Fig. S7. Full linear regression model fit to *SbMATE* expression data for all *SbWRKY1/SbZNF1* haplotypes, which is shown as a linear trend-line in Fig. 4C. *SbWRKY1* and *SbZNF1* effect on *SbMATE* expression was estimated based on RILs homozygous for the SC283 (Al-tolerant) alleles at both TF loci (**Z+/W+**), for the BR007 (Al-sensitive) allele (**Z-/W-**), or showing alternate TF alleles (**Z+/W-** and **Z-/W+**). The regression model was highly significant ($\alpha = 0.01$). The intercept (2.96), the slope (|0.4|) and the coefficient of determination (R^2) are shown on top. The high R^2 value (0.98) indicates that haplotype variation for the TFs largely explain the variation in *SbMATE* expression. 90% confidence intervals for haplotype means are also shown. RQ: relative quantification using the $\Delta\Delta C_t$ method (see methods).

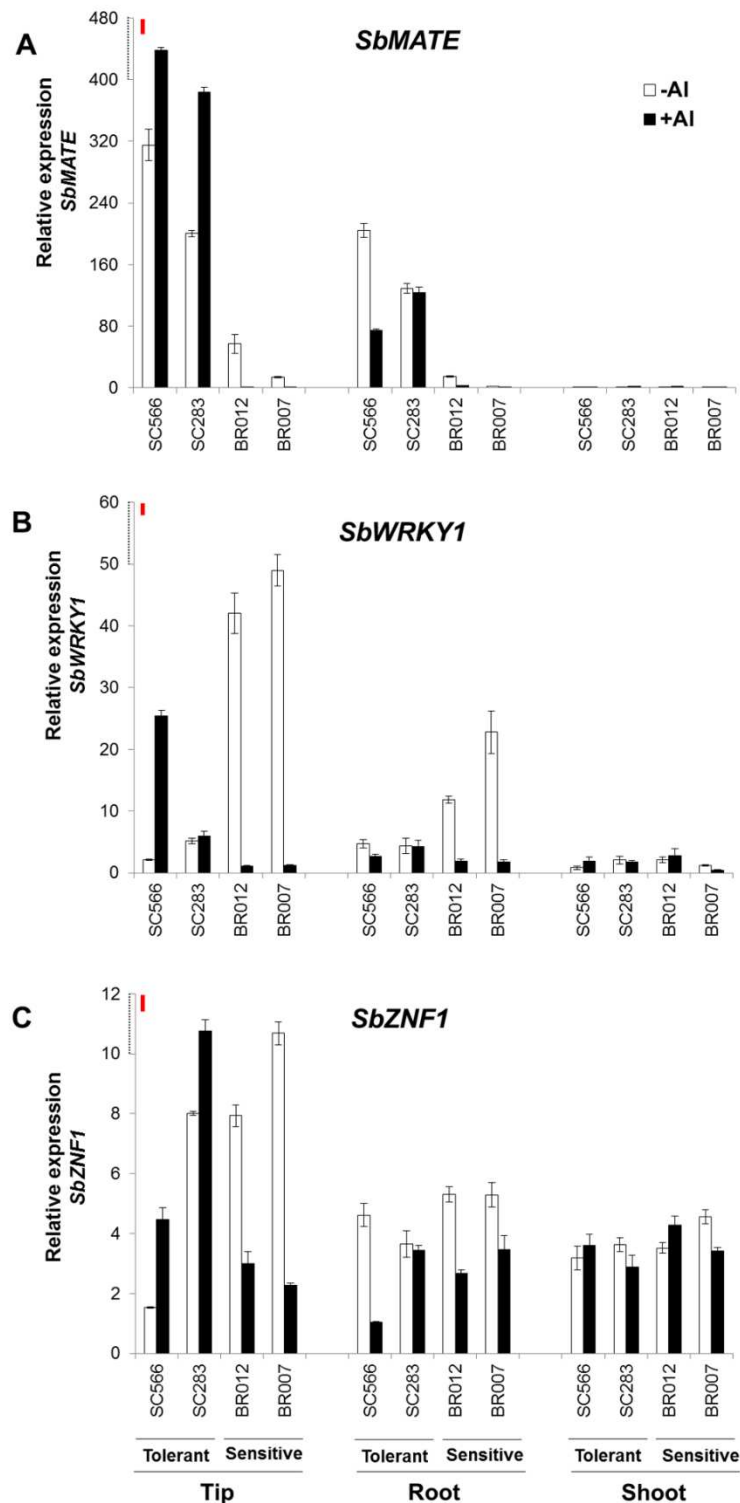


Fig. S8. Tissue-specific expression of *SbMATE* and the interacting transcription factors. (A) *SbMATE*, (B) *SbWRKY1* and (C) *SbZNF1* expression in SC566 and SC283 (Al-tolerant) and the Al-sensitive lines, BR007 and BR012. The sorghum genotypes were grown without (-AI) or with $27 \mu\text{M Al}^{3+}$ (+AI) in nutrient solution at pH 4.0 for 5 days. The root tip (1 cm), a composite sample of the rest of the root system excluding the root apex, and shoots were collected for RNA isolation. Standard deviation bars of three replicates are shown and least significant difference (Fisher's LSD, $\alpha = 0.05$) bars (in red) are drawn to scale for each gene (top of the Y-axis).

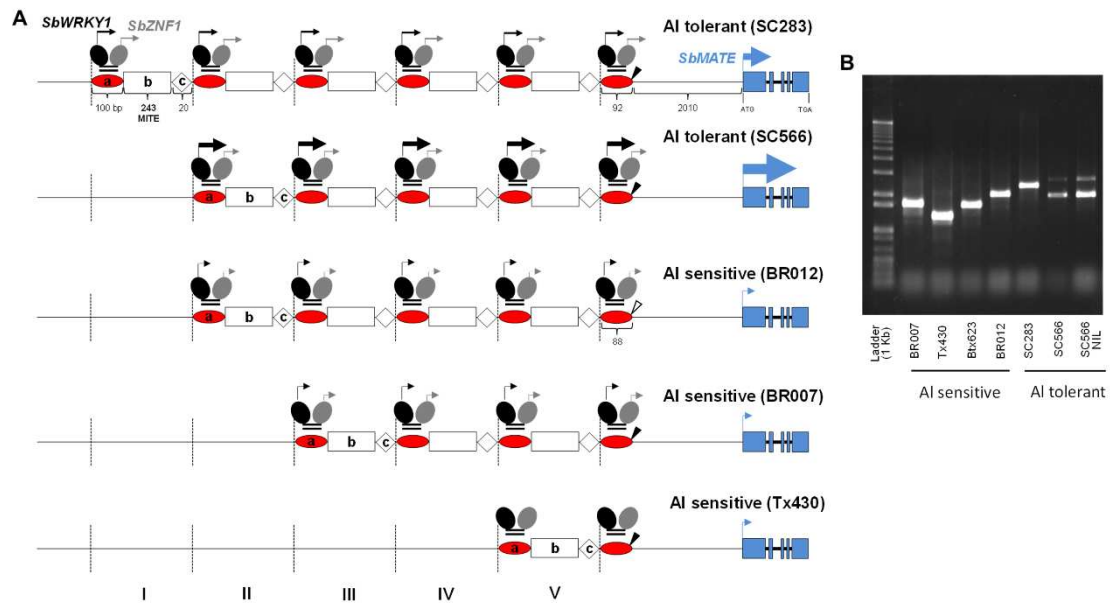


Fig. S9. *SbMATE* is regulated by *cis/trans* interactions. (A) The general structure MITE region is depicted: the 243-bp MITE element (unit “b”) is flanked by 100-bp (unit “a”, 100-bp repeat) and 20-bp (unit “c”) sequences, and the number of identical a-b-c triplets varies in different promoter alleles. The MITE-containing a-b-c triplets (MITE repeats) is terminated by a single 100-bp “a” unit with either an 8-bp deletion (depicted by an inverted black triangle, present in SC283 and Tx430) or a 12-bp deletion (inverted white triangle, present in BR012), which results in the 88- or 92-bp terminal fragments. *SbWRKY1* and *SbZNF1* protein binding is represented by black (*SbWRKY1*) and gray (*SbZNF1*) ovals above the binding fragment (red). The size of the arrows above genes is proportional to relative expression values following a 5 day AI exposure for *SbWRKY1* and *SbZNF1* in Fig. 4A and *SbMATE* in Fig. S6. (B) Amplification profile of the MITE repeats in AI-tolerant and AI-sensitive lines (20). Our model depicts enhanced recruitment of *SbWRKY1* and *SbZNF1* to the *SbMATE* promoter region in AI-tolerant compared to AI-sensitive lines due to increased binding site abundance in the AI-tolerant lines, as well as reduced TF expression in AI-sensitive backgrounds, reducing *SbMATE* expression.

CAPÍTULO II

DIAGNÓSTICO MOLECULAR DO DISTÚRPIO FISIOLÓGICO EM EUCALIPTO

RESUMO

O termo “distúrbio fisiológico” foi cunhado para descrever um estado fisiológico de estresse em eucalipto que tem sido relativamente frequente nas regiões sul e leste da Bahia. Este evento está diretamente relacionado com a perda de dominância apical, levando à necrose do tronco principal, comprometendo deste modo, a produção da cultura e principalmente a qualidade da madeira. Paralelamente, ocorre emergência de um tronco secundário associado a um fenômeno de necrose do caule principal que se torna quebradiço e os sintomas são também associados com perda de turgor, encarquilhamento e amarelecimento das folhas. Os mecanismos moleculares envolvidos no desenvolvimento dos sintomas e na resposta de defesa da planta ao distúrbio são totalmente desconhecidos. Para compreender os pontos-chave deste mecanismo e diagnosticar os eventos moleculares associados ao “distúrbio fisiológico, foi utilizado a tecnologia RNA-Seq para o sequenciamento do transcriptoma a partir de amostras de caules do clone suscetível 520 e do clone resistente 1404. As amostras de caules dos referidos clones foram coletadas no campo, a partir de plantas crescidas no mesmo talhão sob as mesmas condições climáticas e foram assim designadas: 520-S1, que corresponde ao caule principal necrosado ou sintomático, 520-S2, ao caule secundário emergentes, 520-S3 caule assintomático do clone suscetível e 1404-R1, caule do clone resistente e, portanto, assintomático. Este delineamento experimental permitiu inferir que os genes diferencialmente expressos pelo contraste 520-S1 vs 520-S3 representavam o transcriptoma induzido pelo distúrbio fisiológico enquanto que 520-S2 vs 520-S3 representava o transcriptoma induzido em recuperação do estado de estresse e 1404-R1 vs 520-S3 permitiria avaliar um somatório do efeito de resistência e de genótipo. O estabelecimento do distúrbio fisiológico provocou uma reprogramação massiva na expressão gênica, resultando em 4286 genes diferencialmente expressos. Entre os genes regulados positivamente houve um enriquecimento de genes na categoria morte celular e metabolismo secundário ligados à resposta a estresse abiótico.

Diferentemente das amostras S2 vs S3, onde apenas 1311 genes foram diferencialmente expressos, esta diferença nos valores de genes DE faz referência à transição de estado da amostra S2, entre o estado de distúrbio fisiológico e o estado de recuperação, com a emergência de um novo tronco. Por meio de Diagramas de Venn, foi possível analisar as relações entre os conjuntos de genes expressos diferencialmente entre os clones contrastados ao clone controle, e revelaram um total de 60 genes expressos diferencialmente *down* regulados no contraste S1 vs S3 e simultaneamente *up* regulados em R1 vs S3. Em contraposição, foram anotados 26 genes na interseção entre S1 vs S3 e R1 vs S3, *up* regulados no primeiro e *down*

regulados no segundo. Resultados da classificação por Gene Ontology (GO) enriquecidos em relação aos processos biológicos demonstraram ativação dos genes envolvidos em vias de sinalização de morte celular, ativação de processos que envolvem principalmente resposta ao estresse. Em contrapartida, encontraram-se reprimidos genes relacionados a processos de desenvolvimento, replicação do DNA e metabolismo. Juntos, estes dados sugerem que a causa do distúrbio fisiológico provavelmente está relacionada a uma ativação descontrolada da via de morte celular. A presença de genes marcadores selecionados com base em análises de qRT-PCR para validar os dados de RNA-Seq, constituirão a base para desenvolvimento de diagnósticos moleculares em eucalipto, ainda na fase de muda. Posteriormente, análise de possíveis SNPs (*single nucleotide polymorphism*) associados ao distúrbio será realizada com base nos genes marcadores selecionados, a fim de um conhecimento mais apurado sobre as causas moleculares do distúrbio.

1. INTRODUÇÃO

O gênero *Eucalyptus* é originário da Austrália, pertence à família *Myrtaceae*, composto por mais de 700 espécies. Dentro do gênero *Eucalyptus*, são descritos dez subgêneros, sendo *Symphyomyrtus* considerado como principal, com mais de 470 espécies. Entre estas espécies, as mais plantadas para fins comerciais são *E. grandis*, *E. globulus*, *E. urophylla*, *E. calmadulensis*, *E. saligna* e *E. tereticornis* (Eldridge et al., 1994; Brooker, 2000; Poke et al., 2005). A espécie *E. grandis* tem o cultivo localizado em regiões tropicais e subtropicais, enquanto que *E. globulus* concentra-se em regiões temperadas, sendo que seus híbridos são amplamente utilizados para plantações industriais devido à capacidade de adaptação e crescimento rápido (Paiva et al., 2011). As árvores são hermafroditas e protândricas, apresentando fecundação preferencialmente alógama, podendo também se reproduzir por autofecundação, embora haja forte depressão por endogamia (Williams e Woinarski, 1997).

O genoma do eucalipto é diploide, variando entre 600 e 700 milhões de pares de bases, distribuídos em 22 cromossomos ($2n=2x=22$) (Grattapaglia et al., 2012). O sequenciamento completo do genoma do eucalipto (Myburg et al., 2014), iniciado em 2008 pela rede internacional EUCAGEN (Eucalyptus Genome Network), foi liderado por cientistas de três países: Universidade de Pretória (África do Sul); Embrapa Recursos Genéticos e Biotecnologia (Brasil), e Joint Genome Institute – JGI do Departamento de Energia dos Estados Unidos (DOE). O sequenciamento do eucalipto, cujo genoma representa atualmente o chamado "genoma de referência", é da espécie *E. grandis*, o eucalipto tropical mais plantado no Brasil. O estudo analisou 640 milhões de pares de bases do genoma do eucalipto, a decodificação da sequência completa resultou na identificação de todos os 36 mil genes. Desta forma, a compreensão da base molecular da plasticidade do eucalipto representa um elemento chave para a sustentabilidade da indústria de base florestal frente às mudanças climáticas que tendem a aumentar a incidência de secas, geadas e grandes flutuações de temperatura (Grattapaglia et al., 2012).

As espécies do gênero *Eucalyptus* apresentam propriedades físicas e químicas muito diversas, sendo destinadas principalmente à produção de carvão, energia, celulose, estacas, medicamentos, entre outros (Grattapaglia et al., 2012). Assim, este gênero caracteriza-se por apresentar rápido crescimento, elevada produtividade, ampla diversidade de espécies, grande capacidade de adaptação, e pelo seu uso destinado à diferentes finalidades, tem sido amplamente utilizado em plantações florestais (Silveira et al., 2007). Tais motivos, fornecem a base para programas de melhoramento de espécies adaptadas a um amplo espectro de

ambientes, inclusive no Brasil. Vale ressaltar, que as espécies de eucalipto, apesar de seu rápido crescimento, os ciclos de melhoramento geralmente levam de 12 a 16 anos para fornecer genótipos elite (Rezende et al., 2014).

Segundo dados do IBGE, em 2018, o Brasil continha 9,85 milhões de hectares de florestas plantadas, sendo 75,2% de eucalipto e 20,6% de pinus. Em 2016, o Brasil liderou o ranking global de produtividade florestal, com média de 35,7 m³ ha/ano no plantio de eucalipto e 30,5 m³ ha/ano no plantio de pinus, de acordo com os dados da Indústria Brasileira de Árvores (IBÁ). O Brasil, é o 3º maior exportador de celulose, participando com 13,2% do mercado mundial de US\$ 47,98 bi. Do total da produção brasileira 69% destina-se à exportação (Fonte: www.agricultura.gov.br). Os principais plantios estão localizados nos estados de São Paulo 27%, Bahia 26%, Minas Gerais 13% e Espírito Santo 10% (Fonte: www.bracelpa.org.br).

A cultura do eucalipto se encontra distribuída por quase todo o território nacional, sendo que grande parte da área de cultivo apresenta limitações ao desenvolvimento das plantas. Desta forma, uma maior produtividade no setor florestal tem demandado pesquisas à procura de materiais genéticos mais produtivos e adaptados às diferentes condições ambientais. As plantas se encontram frequentemente sob condições desfavoráveis ao seu desenvolvimento (Cambráia, 2005), e estão aptas a sofrer estresses reversíveis ou irreversíveis (Larcher, 2000) causados por fatores bióticos ou abióticos (Taiz e Zeiger, 2004). Fatores ambientais podem ocasionar estresse em uma planta, tais como radiação solar, solo, disponibilidade hídrica e temperatura (Souza et al., 2006).

Em plantações de eucalipto, a disponibilidade de água é um fator limitante para o crescimento e produtividade. O estresse hídrico por escassez de água ocorre quando a baixa disponibilidade de água afeta os processos fisiológicos (Silva, 1998), como transpiração, fotossíntese, abertura estomática (Flexas et al., 2002; Chaves et al., 2003; Wright et al., 2004), produção de ácido abscísico (Taylor et al., 2000; Seki et al., 2007) e ajuste osmótico (Bray, 1997; Callister et al., 2006). Como resposta à deficiência hídrica, as plantas passam por mudanças fundamentais na relação da célula com a água e nos seus processos fisiológicos (Pimentel, 2005) e morfológicos (Chaves et al., 2004), influenciando a sua capacidade de tolerar as condições adversas do meio. Por conseguinte, há restrição ao acúmulo de massa (Silva, 1998), prejudicando tanto no crescimento inicial das plantas como nos estádios mais tardios, limitando a dimensão das folhas individuais, o número de folhas, o número e taxas de crescimento dos ramos e o crescimento do caule (Osório et al., 1998).

Outro fator limitante para o desenvolvimento da cultura, são os solos pobres, muito intemperizados e lixiviados, onde os nutrientes ocorrem em níveis limitantes para o

desenvolvimento das plantas. A aplicação de fertilizantes é necessária para corrigir as deficiências nutricionais desses solos, permitindo dessa forma a obtenção de maior produtividade. No Brasil, o boro é o micronutriente que apresenta níveis mais limitantes nos solos florestais (Silveira et al., 2007). Os fatores que afetam a disponibilidade de boro no solo são caracterizados por pH, textura, umidade, temperatura, matéria orgânica, quantidade e tipos de argila. Os primeiros sintomas de deficiência de boro em eucalipto foram verificados por Vail et al. (1957) e Savory (1962). Esses autores relataram que o principal sintoma é a seca de ponteiro (“dieback”), sendo que as manifestações ocorrem, principalmente, durante as estações secas. Em eucalipto caracterizam-se por folhas novas cloróticas, pequenas e mal formadas, com aspecto de “falta de pedaços”; encarquilhadas e coriáceas; presença de nervuras extremamente salientes com posterior necrose, dando aspecto de “costelamento”; morte da gema apical e seca de ponteiro seguida de brotações das gemas axilares, com posterior bifurcação do tronco; caules e ramos retorcidos devido à falta de lignina; rachaduras da casca com exudação de goma e necrose dos tecidos; morte descendente dos ramos e achatamento do caule devido à morte do câmbio (Tokeshi et al., 1976; Rocha filho et al., 1978; Dell & Malajczuk, 1994; Silveira et al., 1996). No campo, a deficiência severa leva a mortalidade das árvores, enquanto que a deficiência leve pode proporcionar apenas danos nos galhos e troncos (necrose dos tecidos e gomose) e baixa qualidade da madeira. Entretanto deve-se ressaltar que a exigência de boro e os danos causados pela carência diferem entre espécies e genótipos (Stone, 1990). Igualmente ao que ocorre na parte aérea do eucalipto, os sintomas de deficiência de boro manifestam-se também nas raízes.

Além dos efeitos já citados, deve-se ressaltar que a baixa temperatura, também é considerado como um fator abiótico de importância, uma vez que restringe a produtividade e limita áreas de cultivo de plantas. A aclimação ao frio é um processo que induz mudanças na expressão gênica, resultando em alterações fisiológicas e bioquímicas, podendo inclusive causar alterações no teor de lignina das plantas (Moura et al., 2010). Em plantas de *Populus*, observou-se aumento na deposição de lignina após tratamento em temperatura de 10°C (Hausman et al., 2000).

A indução da síntese de lignina, segundo os autores, estaria relacionada ao reforço da parede celular contra estresse mecânico ou por desidratação. Estudos evidenciam, aumento do teor de lignina em *Eucalyptus globulus*, quando submetidos a estresses hídrico e por baixa temperatura (Souza, 2013). A parede celular constituída por celulose, hemicelulose, pectinas, proteínas e lignina, atua na defesa da planta contra ataques de patógenos e estresses abióticos, define o formato da célula, regula o seu crescimento e proporciona suporte mecânico e

estrutural às plantas (Zhong and Ye, 2007). A deposição da lignina na parede celular secundária, após a finalização do alongamento celular é regulado por uma cascata de fatores de transcrição (FTs). Estudos realizados em *Arabidopsis thaliana* têm demonstrado a existência de vários FTs envolvidos na biossíntese de parede celular secundária (McCarthy et al., 2009). Algumas famílias de FTs relacionadas à deposição da lignina são: Família Homeobox (Mele et al., 2003), a família gênica NAC (Nam Ataf CUC2) (Wang et al., 2011; Zhong et al., 2008; Yamaguchi et al., 2011), membros da família LIM (Patzlaff et al., 2003; Kawaoka and Ebinuma, 2001; Kawaoka et al., 2006) e por fim, família MYB (Legay et al., 2007; Jin et al., 2000; Wang and Dixon, 2012; Zhong and Ye, 2009).

Os fatores abióticos não são os únicos responsáveis pela perda da produção do eucalipto, com a expansão da cultura, têm sido relatados um aumento no registro de doenças em viveiros e em plantações florestais. Algumas doenças sem expressão econômica são conhecidas desde a década de 1980, como o cancro e a ferrugem do eucalipto (Krugner, 1980; Ferreira, 1989). O agente causal do cancro do eucalipto é o fungo *Crysoporthe cubensis*. Esta doença foi registrada nas espécies *E. grandis*, *E. oreades* e *E. viminalis*, por Auer (1996). O agente causal da ferrugem é o fungo *Puccinia psidii* Winter, sendo que a doença foi registrada somente em *E. grandis* e *E. dunnii* (Auer et al., 2010). Os principais danos da ferrugem são a redução da área fotossintética e o impedimento do crescimento nos dois primeiros anos de idade das árvores altamente suscetíveis (Ferreira, 1989). Após esta idade, a doença não causa danos. O impacto pode ser verificado na redução da produtividade da floresta nos primeiros anos de crescimento, com conseqüente queda no seu volume final. Além dos fungos, existe um grupo de fitopatógenos da cultura do eucalipto, fitobactérias *Ralstonia solanacearum*, *Erwinia psidii* e *Xanthomonas axonopodis*, também foram relatadas como importantes agentes causais de doenças do eucalipto no Brasil (Alfenas et al., 2009).

Recentemente foi identificado um estado de comprometimento fisiológico no eucalipto, designado “distúrbio fisiológico” associado a um conjunto de sintomas. Entre eles, incluem murcha das folhas, amarelecimento das folhas, seca do ponteiro, multiplicação de ramos na copa devido à perda de dominância apical, necrose do tronco principal na parte inferior devido ou levando à perda de dominância apical, multiplicação de ramos ao longo do tronco e formação de cascas necróticas, enrugadas e secas, ao longo do tronco em árvores adultas. Embora alguns desses sintomas também estejam associados a resposta por patógenos ou resposta à seca, tipicamente, os estágios iniciais do estabelecimento do “distúrbio fisiológico” têm sido associados à perda de dominância apical e ramificação da parte inferior do caule ou da copa, com morte celular e necrose do caule primário, o que o diferencia de outras

anomalias bióticas e abióticas. Assim sendo, torna-se possível delimitar um conjunto de sintomas que resultam nas anomalias de desenvolvimento típicas do “distúrbio fisiológico” e diagnosticar corretamente as plantas acometidas. Claramente, o denominador comum de todos os sintomas, durante a evolução do distúrbio, parece ser a ativação descontrolada de vias de sinalização de morte celular, levando à necrose ou senescência induzida por fatores climáticos.

Morte celular programada (PCD) é uma via de sinalização em plantas ativada por vários estímulos endógenos (como sinais de desenvolvimento) e exógenos (como estresses), sendo um processo altamente dinâmico e regulado, capaz de perceber estresse intracelular dentro de minutos e rapidamente montar uma resposta apropriada para lidar com o dano celular (Batoko et al., 2017). Isto somente é possível porque uma gama de vias de sinalização celulares não relacionadas convergem para ativação da maquinaria de morte celular programada sinalizando para uma diversidade de condições (Reis et al., 2011; Pimenta et al., 2016). De uma forma geral, PCD permite que organismos sésseis como as plantas se protejam contra estresses bióticos e abióticos (Costa et al., 2008; Mendes et al., 2013). Acredita-se que várias vias de sinalização celular de adaptação das plantas ao meio ambiente são, de alguma forma, integradas com a sinalização de morte celular programada (Batoko et al., 2017).

Este conjunto de informações, associados aos sintomas típicos do “distúrbio fisiológico”, é um indicativo forte de que provavelmente vias de sinalização de morte celular e integradas com estresses que permitem as plantas a se adaptarem com o meio ambiente estão sendo inibidas ou ativadas irregularmente nos genótipos suscetíveis em resposta a fatores climáticas. Neste caso, os fenótipos observados são provavelmente resultantes de um somatório dos efeitos genéticos, ambientais e da interação genótipo x ambiente.

Embora tenham sido identificados genótipos resistentes e suscetíveis, as causas da ocorrência do distúrbio fisiológico e os genes associados às respostas moleculares alteradas são totalmente desconhecidos. A elucidação desse fenômeno requererá um conhecimento pleno das vias de sinalização de morte celular ativadas, sua integração com respostas a estresses abióticos, além dos genes e proteínas envolvidas na regulação dessas vias. Devido à abrangência prevista da resposta molecular, este conhecimento somente será possível por meio de estudos moleculares de amplitude genômica. Determinar o transcriptoma induzido pelo distúrbio fisiológico em genótipos suscetíveis e contrastá-lo com genótipos resistentes fornecerá informações valiosas sobre os mecanismos moleculares que levam ao aparecimento da anomalia. O custo e a facilidade em relação ao sequenciamento do genoma têm se tornado uma alternativa economicamente mais viável para a realização do sequenciamento do transcriptoma (Chapelev et al., 2009).

Desta forma, faz-se necessário a interação entre as técnicas biotecnológicas de genômica, transcriptômica, proteômica e bioinformática para permitir a descoberta de novos genes, surgimento de novas perspectivas sobre a estrutura e função de genes e proteínas e a formulação de novas hipóteses sobre o funcionamento de respostas multigênicas complexas (Grattapaglia et al., 2009 e Grattapaglia et al., 2012). O método de estudo do transcrito através de plataformas de sequenciamento de nova geração é conhecido como RNA-Seq (Wang, et al., 2009). Os fragmentos de RNA são geralmente convertidos em cDNA e, em seguida, se ligam a adaptadores em ambas as extremidades. Cada molécula de cDNA é amplificada para ser, então, sequenciada em uma extremidade (single-end) ou em ambas as extremidades (paired-end). O alinhamento das sequências em um genoma de referência permite o mapeamento da estrutura e estimativa dos níveis de expressão de cada gene (Holt & Jones, 2008).

Assim sendo, o objetivo primordial dessa investigação consistiu em determinar e analisar a variação de expressão gênica em genótipos contrastantes para resistência e suscetibilidade ao distúrbio fisiológico. As análises de expressão gênica a partir de RNASeq certamente permitirá a identificação de genes chaves que poderão constituir a base para desenvolvimento de diagnósticos moleculares precoces do distúrbio fisiológico em eucalipto na fase de muda.

2. MATERIAL E MÉTODOS

2.1 Coleta do Material vegetal

Foram utilizados quatro clones de *Eucalyptus grandis*, 0520, 2361, 1404 e 1407 como material contrastante para análise de sequenciamento de RNA. Os clones 0520 e 2361 são classificados como susceptíveis - identificados no campo com sintomas típicos de distúrbio fisiológico, enquanto os clones 1404 e 1407 são resistentes ao distúrbio. Quinze mudas de cada clone, sem sintomas, foram transplantadas para vasos de 2 L de capacidade, com substrato MecPlant® e mantidas em casa de vegetação.

Para coleta dos clones em condições de campo, foram definidas regiões de amostragem no tronco e folhas, assim denominadas: S1- tronco necrosado do clone 0520 e 2361, S2- tronco emergente, S3-controle, oriundo de plantas susceptíveis (mesmo genótipo susceptível) sem sintomas, S4- folhas de plantas susceptíveis com sintomas e S5- folhas de plantas susceptíveis sem sintomas. Quanto aos clones resistentes 1404 e 1407, foram coletadas duas amostras, R1 e R2, similar às amostras S3 e S5, porém referente a um genótipo resistente (Figura 1).

A coleta do material foi feita no campo pertencente à empresa BRASSEL, (Alagoinhas-BA) por técnicos especializados. Foram coletadas três repetições biológicas de cada amostra, plantadas na mesma área de cultivo e sujeitas as mesmas condições climáticas, sob o estágio de desenvolvimento de 4 a 7 meses. Imediatamente após a coleta, ainda em campo, as amostras foram congeladas em nitrogênio líquido, e posterior acondicionadas em gelo seco, para seguir viagem à Viçosa-MG até seu processamento no Laboratório de Biologia Molecular de Plantas/Bioagro/UFV.

2.2 Extração de RNA total das amostras de eucalipto

A extração de RNA total dos clones 0520 e 1404 foi realizada conforme descrito por Chang, Puryear e Cairney (1993) com algumas modificações. Foi utilizado 0,2 g de cada amostra armazenada em -80°C e então macerada em nitrogênio líquido com auxílio de almofariz e pistilo.

Ao material macerado foi adicionado 1 mL do tampão de extração, contendo 2%(w/v) CTAB (cetyltrimethylammonium bromide), 2%(w/v) PVP, 100 mM Tris-HCL, 25 mM EDTA, 20mM NaCl, and 2% β -mercaptoethanol. As amostras foram incubadas por 20 minutos a 65 °C. A esse material adicionou-se 1 mL de clorofórmio, seguida de centrifugação a 15.000 x g,

por 10 minutos, a 4°C. Após uma segunda extração do sobrenadante com igual volume de clorofórmio, RNA total foi precipitado do sobrenadante com igual volume de isopropanol (aproximadamente 400 µL), por 10 minutos a temperatura ambiente e, posteriormente, por centrifugação por 10 minutos a 4 ° C (12.000 x g). Após a centrifugação, o sobrenadante foi descartado, o sedimento foi lavado com 1 mL de etanol a 80% preparado com água DEPC, e os microtubos foram conduzidos para outra centrifugação por 1 minuto a 4 ° C (12.000 x g). O líquido residual foi removido do tubo com uma pipeta e seco sobre papel absorvente, antes de ser ressuspensionado em 20 µl de água tratada com DEPC.

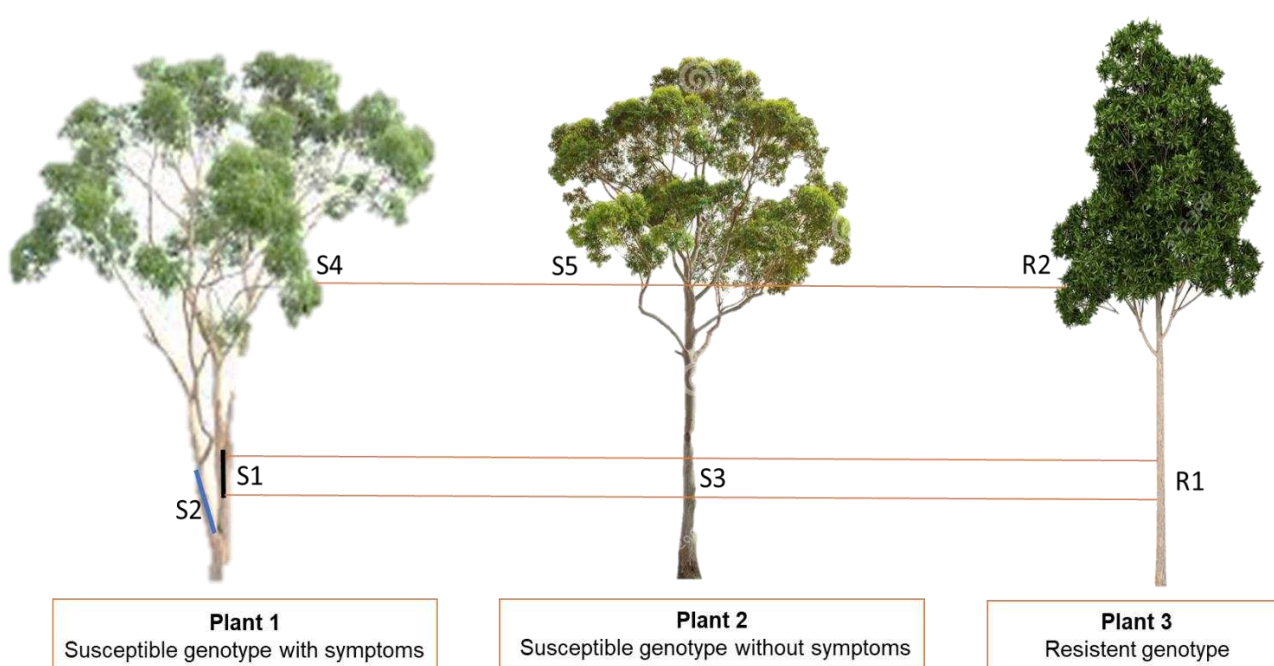


Figura 1. Representação esquemática das amostras de eucalipto.

A planta 1 é representada pelo genótipo suscetível com sintomas, a planta 2 pelo mesmo genótipo, porém (ainda) sem sintomas, ao passo que a planta 3, apresenta genótipo resistente. As amostras S1, S2 e S3 indicam o caule primário necrosado, o caule secundário e o caule primário, na mesma altura dos nós, de plantas sintomáticas, respectivamente. As amostras S4 e S5 indicam as folhas de plantas sintomáticas e assintomáticas, respectivamente. As amostras R1 e R2, são caules e folhas de plantas resistentes. Estas amostras foram coletadas na mesma área de plantio.

2.3 Integridade e preparo das amostras de RNA para envio ao sequenciamento

A quantificação do RNA foi monitorada por espectrofotometria, pela relação de absorvância 260/280 e a integridade destes RNAs foi verificada por eletroforese em gel de agarose desnaturante 1,5 %, e também pelo equipamento 2100 *Bioanalyser* (Agilent Technologies), observando-se a pureza e o tamanho das amostras.

O RNA total de cada uma das amostras foi diluído para 5000 pg/ μ L para a visualização da sua integridade no instrumento 2100 *Bioanalyser* (Agilent). Para isso, foi seguido o protocolo Agilent RNA 6000 pico Assay, de acordo com os procedimentos recomendados com relação à utilização do equipamento e à corrida das amostras, que foi efetuada usando um *chip* específico de RNA, juntamente com os kits de análise de RNA.

O resultado das separações foi registrado para avaliação através do software 2100 Expert que contém um pacote de ferramentas para correções dos eletroferogramas com alinhamento dos picos, cálculo de tamanho e de concentração dos fragmentos.

Como resultados, são fornecidos os valores de RIN (*RNA integrity number*) de cada uma das amostras, que estima a integridade do RNA, juntamente com o eletroferograma dos rRNAs 28S, 18S e 5S, cuja integridade é essencial para a construção da biblioteca de RNAseq. A escala do RIN vai de 0 a 10, considerando o RNA com um número menor que 6 de baixa qualidade.

Desta forma, foram selecionadas inicialmente 4 amostras, com base nos valores obtidos de RIN. As amostras S1.1, S1.2, S1.3; S2.1, S2.2, S2.3; S3.1, S3.2, S3.3 e R1.1, R1.2, R1.3. com três replicatas biológicas, pertencem aos clones 0520 e 1404 que foram enviadas para seu sequenciamento na empresa Macrogen Inc. (Korea).

2.4 Sequenciamento de alto desempenho do transcriptoma – RNA (Seq)

Para o sequenciamento do RNA, utilizou-se a tecnologia da plataforma *Illumina HiSeqTM 2500*, no modo *paired-end*. Cada *flowcell* contém várias faixas com cada faixa produzindo 250 milhões de leituras (ou seja, até 500 GB / *flowcell*) com o comprimento de cada sequência lida de 35 bp a 250 (3402Bioinformatics). O processo de sequenciamento no HiSeq, consiste em etapas: preparo das bibliotecas, geração de clusters e sequenciamento.

2.4.1 Construção da Biblioteca de cDNA.

Antes de proceder à construção da biblioteca de cDNA, a empresa Macrogen avalia a qualidade das amostras usando o Bioanalyzer 2100 da Agilent Technologies mediante o método RNA integrity number (RIN). Amostras com um RIN (próximo ou maior do que 7,0), ou razão perto de 1,5 são recomendadas para prosseguir com a construção da biblioteca.

Para construção da biblioteca de cDNA, a partir das amostras de RNAs totais referidas anteriormente, utilizou-se a metodologia descrita como TruSeq mRNA library construction, de

acordo com protocolo disponibilizado pela Illumina®. As bibliotecas foram preparadas com TrueSeq™ RNA Sample Preparations Kits (ILLUMINA 2011), da qual consiste de cinco etapas principais: (1) seleção de RNAs mensageiros (RNAm) pela cauda poli-A na região terminal 3', usando sondas de oligonucleotídeos de timina (poli-T); (2) formação da cadeia dupla de cDNA; (3) fosforilação da extremidade 5' e reparo das extremidades da cadeia dupla de cDNA; (4) adição da base Adenina à extremidade 3'; (5) ligação dos adaptadores às duas extremidades dos fragmentos de cDNA. Através de um correspondente nucleotídeo T ligado na extremidade 3' do adaptador, o que garante a complementariedade das saliências. Essa estratégia permite hibridização de cDNAs de dentro de uma *flow cell*.

Finalmente os fragmentos que possuem os adaptadores em ambas extremidades são ligados a templates complementares que estão fixos numa superfície sólida chamada célula de fluxo (*flow cell*).

2.4.2 Formação de clusters paired-end

Para a realização do sequenciamento, os *templates* fita simples são primeiramente amplificados para formar clusters clonais na *flow cell* de sequenciamento.

Os adaptadores ligados às extremidades dos fragmentos de cDNA foram hibridizados à *flow cell* pela extremidade 5', na *flow cell* existe uma superfície composta por óligos complementares aos adaptadores das bibliotecas. A extremidade 3' está livre para servir como iniciador na reação de sequenciamento, formando pontes para favorecer a amplificação, na presença da enzima polimerase e de nucleotídeos não marcados. Este processo consiste numa reação de PCR para obter seletivamente mais fragmentos de cDNAs que contêm moléculas de adaptadores em ambas as suas extremidades, e para aumentar a quantidade de *template* da biblioteca. Os ensaios de PCR foram realizados com um coquetel de *primers* complementares às bases das extremidades dos adaptadores. Posteriormente, os amplicons sofrem desnaturação, onde, as fitas são separadas e linearizadas, formando uma nova ponte para permitir a amplificação. Estes passos são repetidos sucessivamente, gerando milhões de *clusters* de um determinado fragmento (Carvalho; Silva, 2010).

4.3 - Sequenciamento

A *flow cell* contendo os clusters já formados é colocada no equipamento HiSeq 2500, onde milhões de *clusters* são sequenciados simultaneamente. Uma nova desnaturação é feita

para dar início a ciclos automatizados de extensão e captura de imagem. Neste instante, ocorre a incorporação dos nucleotídeos marcados com fluoróforos e sequenciamento por síntese base a base em paralelo. Estes nucleotídeos têm propriedades de terminação, o que permite parar a síntese de DNA quando a enzima DNA polimerase integra o nucleotídeo correspondente na fita nascente. Uma vez integrado, os fluoróforos são excitados por uma série de lasers. A luz emitida será diferencial dependendo do nucleotídeo incorporado, e esta informação é captada por câmeras.

Após a incorporação de cada nucleotídeo no fragmento em síntese, ocorre uma etapa de lavagem para remoção dos reagentes excedentes e do fluoróforo do nucleotídeo incorporado no ciclo anterior, para a ligação de uma próxima base na região de sequenciamento (Carvalho; Silva, 2010). O processo se repete por vários ciclos referentes ao sequenciamento da *Read 1*. A fita recém sintetizada foi desnaturada ao final do sequenciamento da *Read 1* e sua extremidade 3' então desbloqueada para síntese da fita molde, permitindo assim, o sequenciamento da *Read 2 (paired end)*.

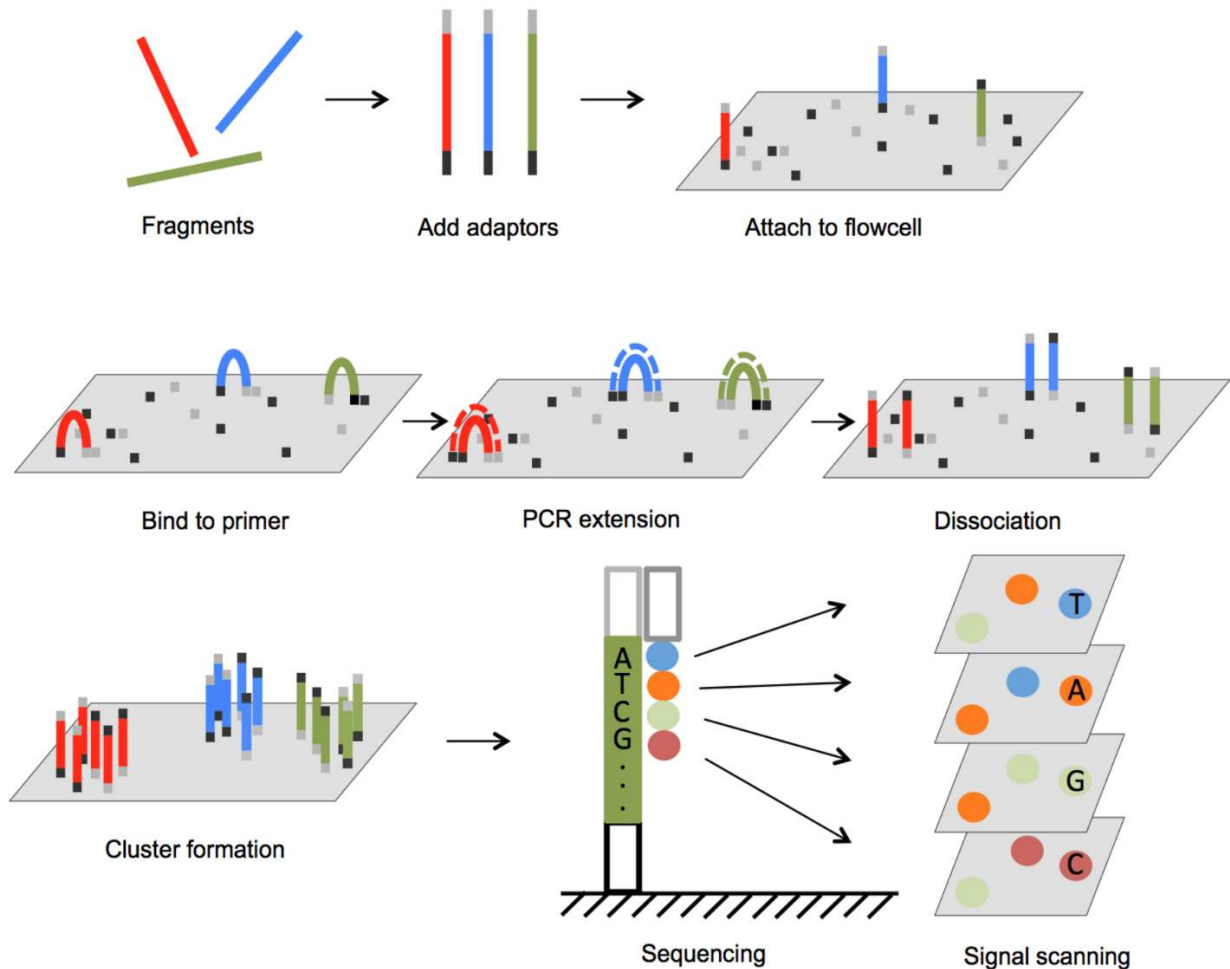


Figura 2. Sequenciamento plataforma *illumina HiSeq™ 2000*.

Os adaptadores são anelados nas extremidades dos fragmentos. Os fragmentos se ligam à *flowcell*, na qual a superfície possui oligonucleotídeos complementares aos adaptadores ligados. Uma reação de PCR amplifica cada fragmento ligado para produzir aglomerados de fragmentos, e agora ambos os fragmentos estão ligados à *flowcell*. Os amplicons sofrem desnaturação, assim, estão livres para uma nova amplificação. Esta PCR é repetida por várias vezes, resultando em um *clusters* de cópias ao redor do fragmento inicial. Inicia-se então o sequenciamento cíclico desses *clusters* de fragmentos. Em cada ciclo de sequenciamento, um nucleotídeo ligado ao fluoróforo é adicionado à fita crescente. Lasers excitam os fluoróforos, e um scanner óptico capta os sinais de cada *clusters* de fragmento. Em seguida, o terminador é removido e o próximo ciclo de sequenciamento é iniciado.

Fonte: disponível em <http://www.3402bioinformaticsgroup.com/service/>

2.5 Bioinformática

Diversas ferramentas de bioinformática foram recentemente desenvolvidas para analisar a imensa quantidade de informação gerada pela técnica de RNA-Seq. Estas ferramentas diferem entre si quanto à normalização e às técnicas estatísticas aplicadas com impacto nos resultados finais (Almeida, 2013).

O sequenciamento de RNA (RNA-Seq) tornou-se uma tecnologia muito utilizada para criar perfis de expressão gênica. A técnica é capaz de revelar um repertório completo de isoformas de *splices* alternativos no transcriptoma e lança luz sobre os transcritos mais raros e específicos das células (Trapnell et al., 2012). Com o perfil de RNA-Seq é possível identificar genes ou vias moleculares que são expressas diferencialmente (DE) entre duas ou mais condições biológicas. Assim, faz-se necessário criar um fluxo de trabalho computacional para a detecção dos genes e vias (Chen et al., 2014).

As primeiras etapas após o processo de sequenciamento, foram realizadas pela empresa MacroGen numa etapa de controle de qualidade. Posteriormente, iniciaram-se as análises com ferramentas de bioinformática, conforme esquematizado na Figura 3.

2.5.1 Verificação da qualidade das leituras geradas

A empresa MacroGen realizou a primeira etapa de análises após o sequenciamento: O *software* de conversão Illumina bcl2fastq2 foi utilizado para conversão dos dados gerados no sequenciamento para o formato de sequência *fastq*.

Posteriormente, foi realizado o controle de qualidade das *reads*. As *reads* representam sequências curtas provenientes do sequenciamento (em número de milhões), para isso utilizou-se a ferramenta *FastQC*. Este software analisa parâmetros como a qualidade da sequência por nucleotídeo (Q20 e Q30), por conteúdo de GC, por conteúdo de nucleotídeos não determinados (N) e outros. *FastQC* usa o *phred quality score* para determinar os Q20 e Q30 score, a partir da probabilidade de que uma base na leitura tenha sido atribuída de modo errôneo durante o sequenciamento.

E então, foi feita a remoção dos adaptadores Illumina, utilizados para ancoragem dos *primers* de sequenciamento. A remoção dos adaptadores é de grande importância, devido ao fato que, podem ser incorporadas na montagem dos genes ou no mapeamento do genoma de referência, contribuindo, desta forma, para uma montagem errônea das sequências. As sequências dos adaptadores, foram eliminadas com o software *Trimmomatic* (Bolger et al., 2014). Na trimagem dos dados, para remoção das leituras e sequências de baixa qualidade, foi utilizado o *FASTX-Toolkit* (http://hannonlab.cshl.edu/fastx_toolkit/). A análise de qualidade das sequências foi realizada antes e após a limpeza das *reads*, pelo programa *FastQC*.

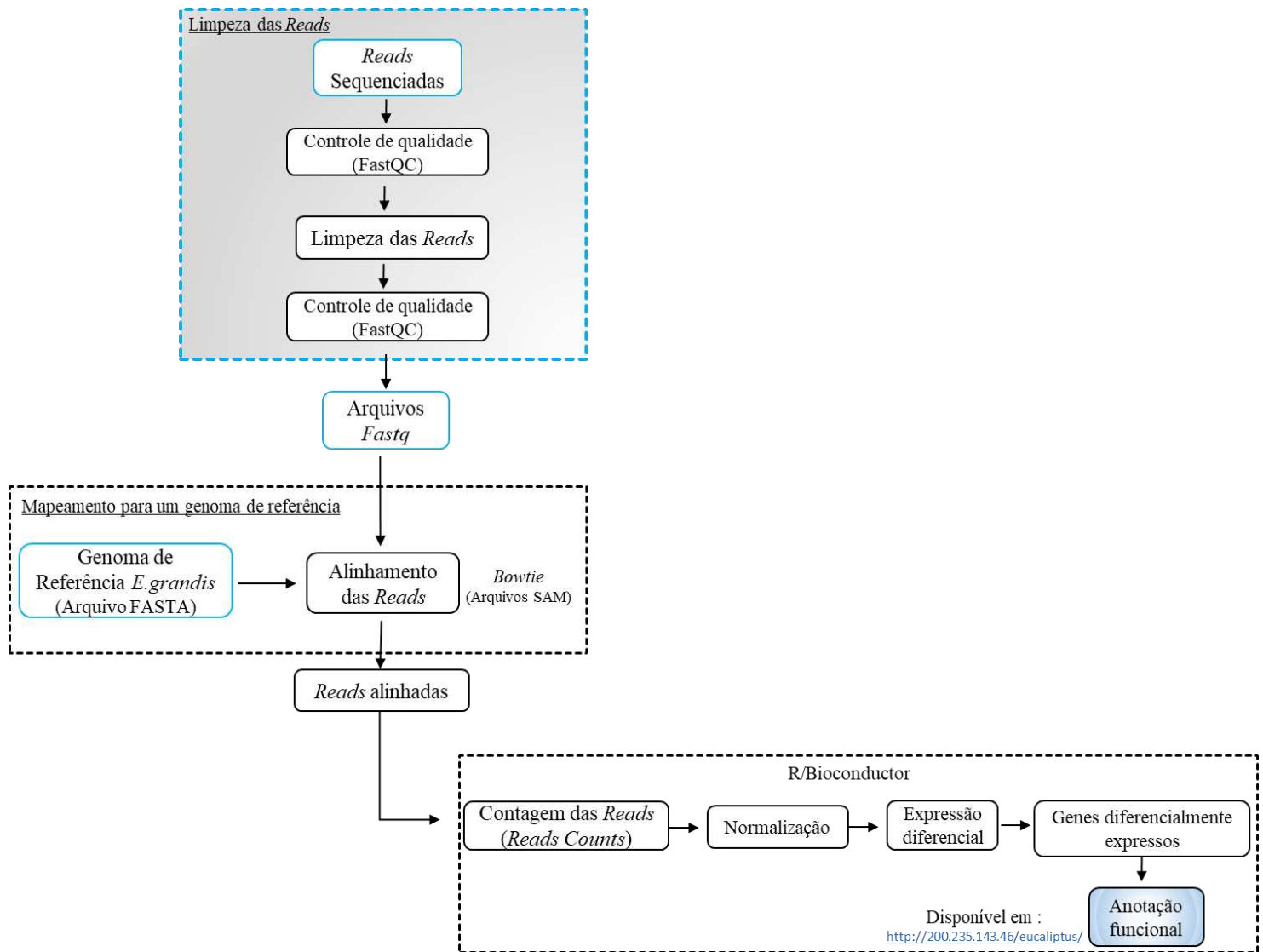


Figura 3. Representação esquemática das diferentes ferramentas de bioinformática envolvidas nas análises de RNA-Seq.

Fonte: Adaptado de (J. Soewarto et al., 2018).

2.5.2 Alinhamento das leituras, montagem de transcritos, análises de expressão diferencial dos genes e mapeamento das redes de co-expressão e modelos de sinalização celular alterados.

O alinhamento das *reads* no genoma de referência de *E. grandis* (versão 2.0) (Myburg et al., 2014) disponibilizado no portal *Phytozome* (<https://phytozome.jgi.doe.gov/pz/portal.html>), foi realizado utilizando o *software* HISAT2 versão 2.05 (Kim et al. 2015). O *software* é específico para alinhamento de sequências curtas de grandes genomas, com maior eficiência e acurácia, de uma maneira rápida e consumindo pouca memória devido a forma como este indexa os dados.

Sabendo-se que o número de *reads* mapeados para cada tipo de RNA é linearmente correlacionado com a sua abundância dentro da célula (número de *reads* α nível de expressão do gene), a próxima etapa a partir do arquivo resultante do alinhamento, foi reunir as *reads* numa unidade com algum significado biológico. A contagem das *reads* mapeadas por gene foi feita pela função *featureCounts* implementada no pacote Rsubread (Liao et al., 2013), disponível no programa R/Bioconductor.

Essas contagens precisam ser normalizadas para a remoção de viés técnicos decorrentes dos passos necessários para análise do RNA-Seq, portanto, é preciso equacionar os dados para uma medida mais correta da taxa de expressão dos RNAs. Para a normalização dos *counts* foram utilizados pacotes disponíveis no Bioconductor.

Para análise de expressão gênica diferencial (DGE) utilizou-se o pacote R/Bioconductor DESeq2 (Love et al, 2014) onde o valor “p” foi corrigido pela taxa de descoberta falsa. O valor de p significativo foi ajustado para 0,05. A análise de expressão diferencial, utiliza o número de *reads* reportadas para cada gene em cada amostra. Diversos modelos estatísticos baseados no teste exato da binomial negativa e nos modelos lineares generalizados foram testados com base nos resultados de contagem das *reads*, para então obter os genes diferencialmente expressos (Anders et al, 2013).

Posteriormente, novas regiões do genoma onde ocorrem expressão foram identificadas por meio do programa Genome Browser e, portanto, novos genes foram estudados. Análises de expressão diferencial de éxons foram estudadas usando o pacote DEXSeq (um pacote do Bioconductor) para a detecção de novas formas alternativas de “*splicing*” (Anders et al, 2012). Esses dados foram anotados usando o banco de dados Phytozome 12.1 (*Eucalyptus grandis* v2.0) e armazenados em um banco de dados relacional, PostgreSQL 10, e uma interface de

acesso foi implementada em PHP. Este sistema está disponível *online* e presente no servidor do Laboratório de Biologia Molecular de Plantas, Bioagro/UFV.

Para o refinamento das análises, primeiramente, foi feita uma análise de enriquecimento gênico usando o teste da exato de Fisher para categorias da ontologia gênica (GO), pelo pacote GStats, do R / Bioconductor (Falcon e Gentleman, 2007) e das vias metabólicas presentes na base dados “Plant Metabolic Network” (Sheth et al, 2014). Por meio dos ortólogos identificados, rede de interação proteína-proteína e de interações genéticas foram construídas. Usando a biblioteca de construção de grafos Boost Graph Library (BGL) em C++ foi construída *in silico* uma rede dinâmica onde diversas operações tais como centralidades, caminhos mais próximos e caminhos de máxima significância poderão ser usados para inferir novas vias de sinalização e regulação.

A expressão diferencial de genes considerados potencialmente relevantes por meio da análise de biologia sistêmica foi validada por meio de RT-PCR quantitativo.

2.6 Extração de RNA dos clones de eucalipto e síntese de cDNA

A extração de RNA dos clones 0520, 2361, 1404 e 1407 foi realizada pelo método, CTAB, seguido por precipitação com isopropanol, conforme descrito no subitem 2.0. A integridade do RNA foi avaliada por eletroforese em gel de agarose 1,5%. O RNA foi quantificado no espectrofotômetro e a concentração foi expressa em $\mu\text{g} \times \mu\text{L}^{-1}$. O RNA total isolado (4 μg) foi tratado com DNase livre de RNase (Invitrogen) e utilizados para obtenção do cDNA, usando a enzima transcriptase reversa M-MLV (Invitrogen) e oligo-dT, de acordo com determinação do fabricante.

2.7 Análises de transcritos por qRT-PCR

Os *primers* utilizados para detectar o acúmulo de transcritos com RT-PCR em tempo real, foram projetados usando o Primer Express 3.0 (Life Technologies), com base nos genes selecionados. A seleção dos 14 genes de interesse, foi feita a partir dos dados de contrastes entre clone 1404 R1 vs clone 520 S3, clone 1404 R1 vs clone 520 S1, clone 1404 R1 vs clone 520 S2, clone 520 S1 vs clone 520 S2, clone 520 S1 vs clone 520 S3 e clone 520 S2 vs clone 520 S3 disponível em <http://200.235.143.46/eucaliptus/>, com base nos valores de *Log₂ Fold Change* e *p* valor . Os *primers* estão listadas na Tabela 1.

Anterior às análises de qRT-PCR, foram realizados testes para verificar a eficiência de amplificação dos *primers* desenhados, sendo utilizadas as condições dos *primers* com eficiência maior ou igual à 95 %. As reações de RT-PCR em tempo real foram realizadas em um equipamento 7500 Real Time PCR System (Applied Biosystems) utilizando a metodologia *SYBR Green PCR Master Mix* (Life Technologies). Como referência endógena, foram testados os genes *eEF2* (Eukaryotic translation elongation factor 2), *ACT* (Actin-related protein 3) e *30S* (30S ribosomal protein S5, chloroplastic).

As reações de qRT-PCR foram preparadas em um volume final de 10 μ L, contendo template de cDNA, *primers* específicos (*forward* e *reverse*) e corante Fast *SYBR Green PCR 2x Master mix* (Applied Biosystems). As condições de termociclagem foram: 95 °C por 10 minutos e 40 ciclos de desnaturação a 94°C por 15 segundos seguido de anelamento e extensão a 60 °C por 1 minuto. A especificidade dos produtos amplificados foi analisada pelas curvas de dissociação geradas pelo equipamento. A expressão gênica foi quantificada utilizando o método comparativo de Ct: $2^{-\Delta Ct}$.

Tabela 1. Lista de *primers* usados para análises de qRT- PCR.

Primer	Eucalipto	Descrição	Sequência (5'-->3')
qRT-EG-KASIII-Fwd	Eucgr.E00458	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III	GCATCGACTGGCTGTTGCTT
qRT-EG-KASIII-Rvs	Eucgr.E00458		CTCCCGAGGGATTTCCAAAC
qRT-EG-ACC1-Fwd	Eucgr.F00484	acetyl-CoA carboxylase 1, synthesis of malonyl-CoA	CCTCGTGCTGCCATTCGT
qRT-EG-ACC1-Rvs	Eucgr.F00484		TCAAGTGCTCGATGAAGCTATCC
qRT-EG-KCS1-Fwd	Eucgr.H00232	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C Term	TTGGCCTTTAAGCATTTTTGC
qRT-EG-KCS1-Rvs	Eucgr.H00232		TTGGTGAGCTCGAGGCTCTT
qRT-EG-ADH-Fwd	Eucgr.K00927	GroES-like zinc-binding dehydrogenase family protein	CACGCATTCTGAAGTCCATCTC
qRT-EG-ADH-Rvs	Eucgr.K00927		CGCTGCACTCGAAGCAGTAG
qRT-EG-PGL1-Fwd	Eucgr.K00013	NagB/RpiA/CoA transferase-like superfamily protein	GGTCATGTGGCGTCTCTGTTC
qRT-EG-PGL1-Rvs	Eucgr.K00013		GGCGAGTCCTTGATGAAAGTG
qRT-EG-RSW10-Fwd	Eucgr.D01028	NagB/RpiA/CoA transferase-like superfamily protein	GCAAGTGAAGCAGCAAATCAG
qRT-EG-RSW10-Rvs	Eucgr.D01028		GCAAGTGAAGCAGCAAATCAG
qRT-EG- Glicosilase- Fwd	Eucgr.B03112	DNA glycosylase superfamily protein	CGGATCGTTCAGCAGCTACA
qRT-EG-Glicosilase- Rvs	Eucgr.B03112		GGACGTTCCCTCGGGTACTTG
qRT-EG-TIL1-Fwd	Eucgr.B03882	POL2A,TIL1; DNA polymerase epsilon catalytic subunit	CAGACATGAACATCACGGACCTA
qRT-EG-TIL1-Rvs	Eucgr.B03882		GTGCTCTTATACCAGCTTTGATTGAC
qRT-EG-MSH2-Fwd	Eucgr.B00700	MSH2; DNA mismatch repair protein MSH2	ACCGATAGATCGTTGATAATCATTGA
qRT-EG-MSH2-Rvs	Eucgr.B00700		ATGGCCCATGCTAAACCAAA
qRT-EG-BARD1-Fwd	Eucgr.D00242	ATBRCA1,BRCA1; breast cancer susceptibility1	CAGCCTTGAGATGCCAGACA
qRT-EG-BARD1-Rvs	Eucgr.D00242		CGAGGGCCTCTGCACTAGAC
qRT-EG-GALACTOSE OXIDASE-Fwd	Eucgr.B03483	Galactose oxidase/kelch repeat superfamily protein	CTGTGGAAGGCCTGTTTTGG
qRT-EG-GALACTOSE OXIDASE-Rvs	Eucgr.B03483		GATCCACCCGCCACTGAATA
qRT-EG-RPT2-Fwd	Eucgr.B03325	Root phototropism protein 2	AGTGGGCAGAACCGAATTTG
qRT-EG-RPT2 -Rvs	Eucgr.B03325		TGGATCCATGACGCTGCAT
qRT-EG-GAPDH-Fwd	Eucgr.J02288	Glyceraldehyde 3-phosphate dehydrogenase	TTGGCATTGTTGAGGGTCTA
qRT-EG-GAPDH -Rvs	Eucgr.J02288		AAGCAGCTCTCCACCTCTC
qRT-EG-UBQ-Fwd	Eucgr.K00756	Small ubiquitin-like modifier 2	GAGGGACATCTATCTCTATGAC
qRT-EG-UBQ -Rvs	Eucgr.K00756		CAACAGTAAGCACACGAG
qRT-EG-eEF2-Fwd	Eucgr.B02912	Eukaryotic translation elongation factor 2	TGAGGTTAATGGGATCGTG
qRT-EG- eEF2 -Rvs	Eucgr.B02912		CAGATTATCGTCATCGACCA
qRT-EG-Actin-Fwd	Eucgr.H02449	Actin-related protein 3	TAAGCATGACAAGGAACCAG
qRT-EG- Actin -Rvs	Eucgr.H02449		TCAGGTCCAAGAAATCGT
qRT-EG-30S-Fwd	Eucgr.D00609	30S ribosomal protein	GGTGTGAGAATGCTTTGG
qRT-EG-30S -Rvs	Eucgr.D00609		TCACTTCCACAATTCTTCC

3. RESULTADOS

3.1 Qualidade das amostras selecionadas para o sequenciamento

Embora o distúrbio fisiológico tenha se tornado um desafio para a qualidade da madeira de Eucalipto, não se sabe quais os fatores ambientais que induzem este estado fisiológico anormal e, portanto, ensaios controlados em casas de vegetação para indução de sintomas não são possíveis. Com base na existência de genótipos suscetíveis e resistentes ao distúrbio fisiológico presentes em campo, cultivados em condições climáticas e em estádios de desenvolvimento similares, foi possível realizar a análise do transcriptoma induzido pelo distúrbio fisiológico, como forma de diagnóstico pela identificação de possíveis genes diferencialmente expressos em genótipos contrastantes (suscetíveis assintomáticas vs suscetíveis sintomáticas e resistentes vs suscetíveis).

Amostras do clone 520 (suscetível) de eucalipto, de plantas em estado de acometimento, designadas S1 (caule necrosado), S4 (folhas sintomáticas) e S2 (caule emergente), de plantas assintomáticas, caules (S3) e folhas (S5), além de amostras do clone 1404 (resistentes), designadas R1(caule) e R2 (folhas) foram utilizadas para extração do RNA total. A qualidade do RNA total foi avaliada, utilizando Bioanalyzer 2100 da Agilent Technologies, antes de proceder ao envio das amostras para sequenciamento. A qualidade das amostras foi determinada mediante o método *RNA integrity number* (RIN) que estima a integridade do RNA. As amostras que obtiveram um RIN próximo ou maior do que 7,0 seguiram para sequenciamento. Desta forma, apenas S1, S2, S3 e R1(amostras de caule) prosseguiram, ao passo que, o RNA extraído de folhas (S4, S5 e R2) não passou no controle de qualidade do Bioanalyser 2100 e, portanto, não foram enviados para o sequenciamento (Tabela 2).

Tabela 2. Controle de qualidade do RNA.

Clones	Amostras	RIN
Clone 520 S1.1	Caule	7.0.
Clone 520 S1.2	Caule	8.1
Clone 520 S1.3	Caule	8.0
Clone 520 S2.1	Caule	8.2
Clone 520 S2.2	Caule	7.3
Clone 520 S2.3	Caule	7.6
Clone 520 S3.1	Caule	8.6
Clone 520 S3.2	Caule	7.6
Clone 520 S3.3	Caule	7.5
Clone 520 S4.1	Folha	5.0
Clone 520 S4.2	Folha	4.73
Clone 520 S4.3	Folha	5.7
Clone 520 S5.1	Folha	6.5
Clone 520 S5.2	Folha	5.3
Clone 520 S5.3	Folha	4.1
Clone 1404 R1.1	Caule	7.8
Clone 1404 R1.2	Caule	7.0
Clone 1404 R1.3	Caule	7.8
Clone 1404 R2.1	Folha	6.3
Clone 1404 R2.2	Folha	6.0
Clone 1404 R2.3	Folha	5.7

3.2. Avaliação de qualidade de leitura geradas pelo sequenciamento, com FastQC, TRIMAGEM e limpeza das sequências de Reads

De acordo com os resultados do sequenciamento, em relação a qualidade das leituras geradas, foram obtidos gráficos através do programa FastQC, por ferramentas de bioinformática, indicando os valores de qualidade de cada base no formato FastQ. Três parâmetros são mostrados, Q20, Q30 e conteúdo de GC% (Figura 4A). Q20 nos indica a porcentagem das leituras com 99% ou mais de acurácia, ou seja, a probabilidade de que uma vez em 100 vezes uma base seja atribuída de forma errada nos picos do cromatograma durante o sequenciamento. O resultado de Q30 representa 99.9% ou mais de acurácia. Quando um sequenciamento alcança um *phred quality* Q30, praticamente todas as leituras serão perfeitas, com zero erros e ambiguidades (ILLUMINA, 2011). As cores de fundo representam a qualidade da amostra, sendo verde muito boa, laranja mediana e vermelho pobre em

qualidade. Assim, o eixo Y representa o índice de qualidade, enquanto o eixo X indica a posição em *read* (bp) (Figura 4A).

As figuras correspondentes às amostras S1, S2, S3 (clone 520) e R1 (clone 1404), demonstram alta qualidade *Phred quality score* (índice de qualidade por base) (Figura 4A). Mesmo assim, foi realizado a limpeza dos dados do sequenciamento, devido a presença de adaptadores nas sequências, de trechos que foram mal sequenciados e que apresentam baixa qualidade (Figura 4B e 4C). Para isso, foi efetuada a remoção destes adaptadores, trimagem e limpeza das sequências de *reads* e novos gráficos foram gerados *Per base quality*, tornando evidente a boa qualidade das amostras conforme Figura 5; o que indica um alto grau de confiança nas sequências geradas.

3.3 Transcriptoma induzido pelo distúrbio fisiológico

A amostra 520-S3, oriunda de plantas suscetíveis (clone 520) sem sintomas (caule susceptível assintomático – amostra S3), foi utilizada como controle de referência, e desta forma, nos permitiu examinar a variação global na expressão gênica dos genótipos contrastantes (caule necrosado, sintomático – amostra S1, caule emergente, sintomático – amostra S2 e caule resistente – amostra R1) associados ao estabelecimento do distúrbio fisiológico (Figura 6). Usando um p-valor de 0,05 e $\log_2\text{-fold-change} > 1,0$ para genes *up* regulados e valores de $\log_2\text{-fold-change} < -1,0$ para genes *down* regulados, foi observado que, o estado de distúrbio fisiológico (520-S1 vs 520-S3) exibiu 4286 genes diferencialmente expressos (DE), dos quais 62,2 % (2666) foram *down* regulados nas amostras S1 em relação à S3, e 37,8% (1620) *up* regulados, diferentemente do contraste 520-S2 vs 520-S3, que apresentou apenas 1131 genes expressos diferencialmente, sendo 58% (757) *up* regulados e 42% (554) *down* regulados (figura 6). Esta alta divergência entre os genes diferencialmente expressos no clone 520-S1 (tronco necrosado, sintomático) quando comparado ao clone 520-S2 (tronco emergente sintomático) evidencia que a amostra S2 representa um estágio de transição entre o estado de distúrbio fisiológico (responsável por uma expressão gênica massiva no caule necrosado) e o estado de recuperação (controle –S3) com a emergência de novo tronco.

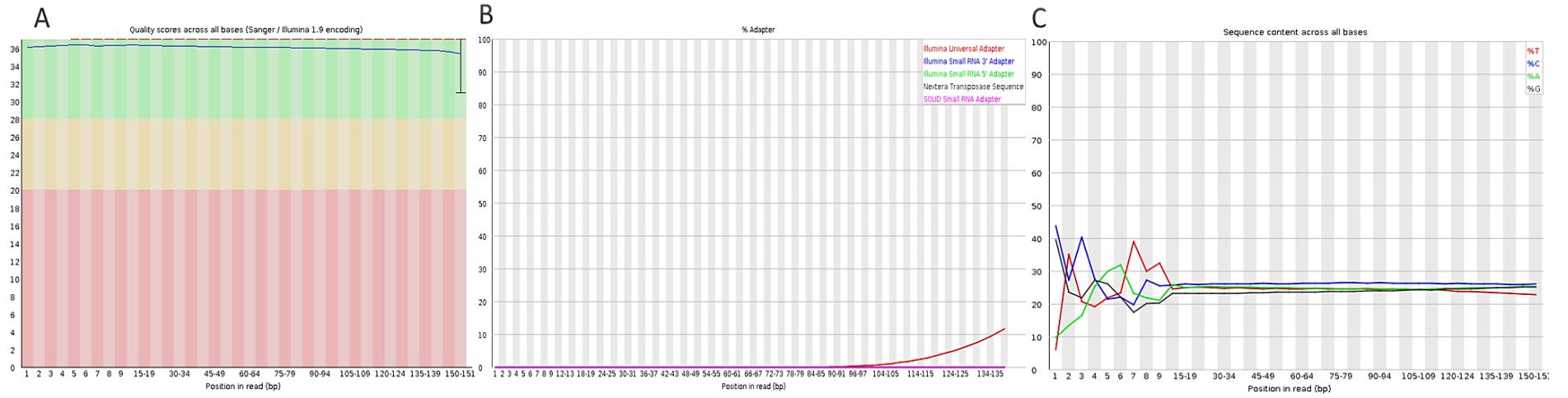


Figura 4: Gráficos *Per base quality* obtido através do FastQC. A) *Phred quality score*, evidenciando boa qualidade das sequências geradas. B) Porcentagens dos adaptadores em relação à posição nas *Reads* em bp. C) conteúdo da sequência por bases.

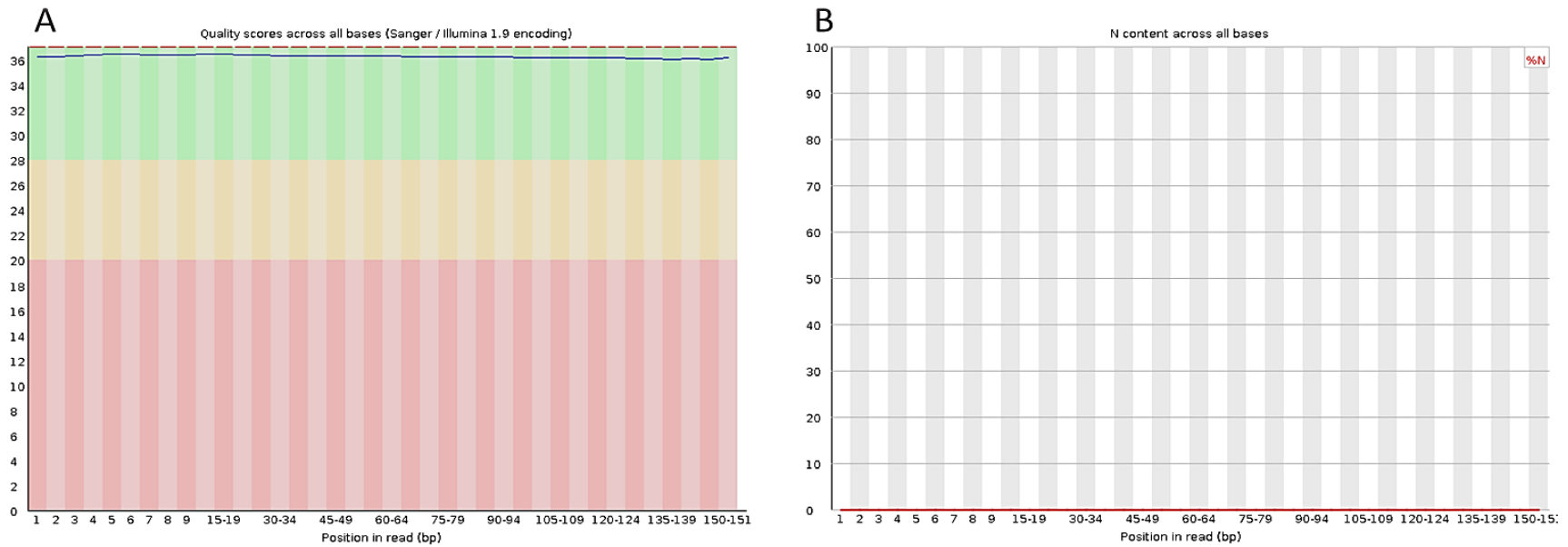


Figura 5: A) Gráfico *Per base quality* obtido através do FastQC, após remoção de adaptadores (B), trimagem e limpeza das sequências de *reads*.

Uma alta porcentagem de genes *down* regulados presentes nas amostras 520-S1 quando contrastado às amostras controle, 520-S3 (plantas suscetíveis sem sintomas), indica que o distúrbio fisiológico induziu uma rede complexa de respostas adaptativas no caule do clone suscetível, tal fato corrobora com o fenótipo encontrado no campo. Uma vez que estas amostras são provenientes de clones de mesmo genótipo, em estágio de desenvolvimento similar, sob as mesmas condições ambientais de crescimento, os genes diferencialmente expressos nas amostras sintomáticas-S1 em relação aos da amostra assintomáticas-S3 provavelmente representam o transcriptoma induzido pelo distúrbio fisiológico. Consequentemente, os genes diferencialmente expressos pelo contraste S1 vs S3 estão relacionados com os sintomas responsáveis pelo distúrbio.

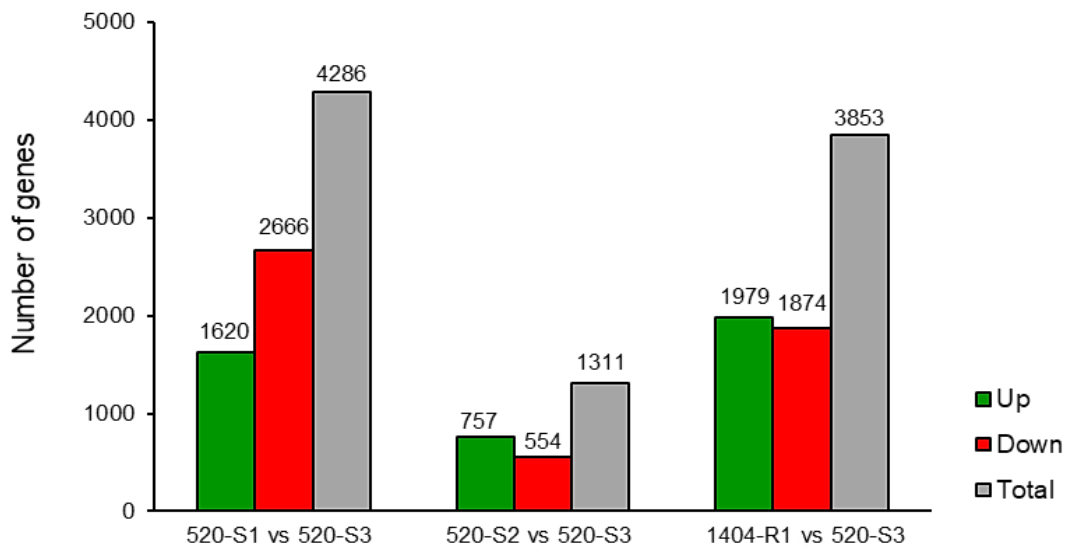


Figura 6: Variação global na expressão gênica induzida pelo distúrbio fisiológico em eucalipto. O eixo y, indica o número de genes diferencialmente expressos. O eixo x, indica contrastes entre os clones 520-S1 vs 520-S3, 520-S2 vs 520-S3 e 1404-R1 vs 520-S3. A análise foi realizada em troncos sintomáticos e assintomáticos do clone suscetível 0520 e do clone resistente 1404.

Por outro lado, a variação global da expressão gênica observada pelo contraste 104-R1 vs 520-S3 deve refletir as mudanças no transcriptoma resultante do *background* genético dos genótipos, da resposta diferenciada dos referidos genótipos às condições ambientais de crescimento e da condição de resistência ao distúrbio fisiológico. Uma comparação das mudanças no transcriptoma dos clones 1404-R1 vs 520-S3 identificou 51,36% (1979) de

genes regulados positivamente (*up*) e 48,63 % (1874) de genes com regulação negativa (*down*), totalizando 3853 genes diferencialmente expressos no referido contraste. Estes genes DE possivelmente compõem o transcriptoma induzido pelo estado de resistência ao distúrbio fisiológico e o transcriptoma diferenciado entre os dois genótipos (1404-genótipo resistente, 520-genótipo suscetível), ou seja, os genes DE devem representar um somatório entre efeito da resistência + efeito do genótipo.

A princípio, o distúrbio fisiológico tem sido associado à morte celular induzida por fatores climáticos, uma vez que os caules se apresentam com sinais visíveis de necrose, quebradiços e as folhas, com clorose, amarelecimento e sinais de murcha. A ativação descontrolada de vias de sinalização de morte celular desencadeada por estresses múltiplos provoca uma série de alterações fisiológicas, metabólicas homeostáticas e de desenvolvimento. Desta forma, para facilitar a interpretação dos resultados e análise dos genes associados às respostas moleculares alteradas, foram categorizados funcionalmente os genes diferencialmente expressos via Gene Ontology (GO) (<http://geneontology.org>), utilizando os critérios $\log_2\text{-fold-change} > 1,0$ para genes *up* regulados, $\log_2\text{ fold-change} < -1,0$ para genes *down* regulados e p-valor de 0,05. Os resultados para classe “processos biológicos” então, representados de forma gráfica nas Figuras 7,8 e 9.

O transcriptoma induzido pelo distúrbio fisiológico (520-S1 vs 520-S3) deve representar a sintomatologia do fenômeno. As categorias biológicas mais enriquecidas de genes DE com expressão diferencial positiva (*up* regulados), foram processo de oxidação-redução (16.7%), transportadores transmembrana (8%) e regulação de transcrição (7.4%) (Figura 7).

Os genes envolvidos em processos de oxidação-redução fazem parte do sistema antioxidativo de defesa da planta, normalmente regulados positivamente em situações de estresses que produzem espécies reativas de oxigênio (ROS) (Sharma et al., 2012). Exemplos nesta categoria regulados positivamente pelo estado de distúrbio incluem os genes da superfamília de peroxidases (Eucgr.A01166, Eucgr.A01166, Eucgr.I01176, Eucgr.L02460); superoxide dismutase (Eucgr.E00505, Eucgr.F01775, Eucgr.F01776, Eucgr.H04426, Eucgr.I01176, Eucgr.I02673); peroxidase (Eucgr.A01166, Eucgr.I01176, Eucgr.I02673, Eucgr.L02460); catalase (Eucgr.F01775, Eucgr.F01776), e desidrogenases (Eucgr.A00387,

Eucgr.A00388, Eucgr.B02479, Eucgr.B03349, Eucgr.C01003, Eucgr.C03047,
 Eucgr.C03157, Eucgr.C03857, Eucgr.C03858, Eucgr.D00468, Eucgr.D00471,
 Eucgr.D01088, Eucgr.E00181, Eucgr.E01103, Eucgr.E01104, Eucgr.E01107,
 Eucgr.E01110, Eucgr.E01117, Eucgr.E01119, Eucgr.E01364, Eucgr.F02744, Eucgr.F04402,
 Eucgr.G02223, Eucgr.H02433, Eucgr.H03936, Eucgr.I00009, Eucgr.I01803, Eucgr.J00610,
 Eucgr.J02207, Eucgr.K00212, Eucgr.K02121, Eucgr.K02125, Eucgr.K02858,
 Eucgr.L01277, Eucgr.L01765).

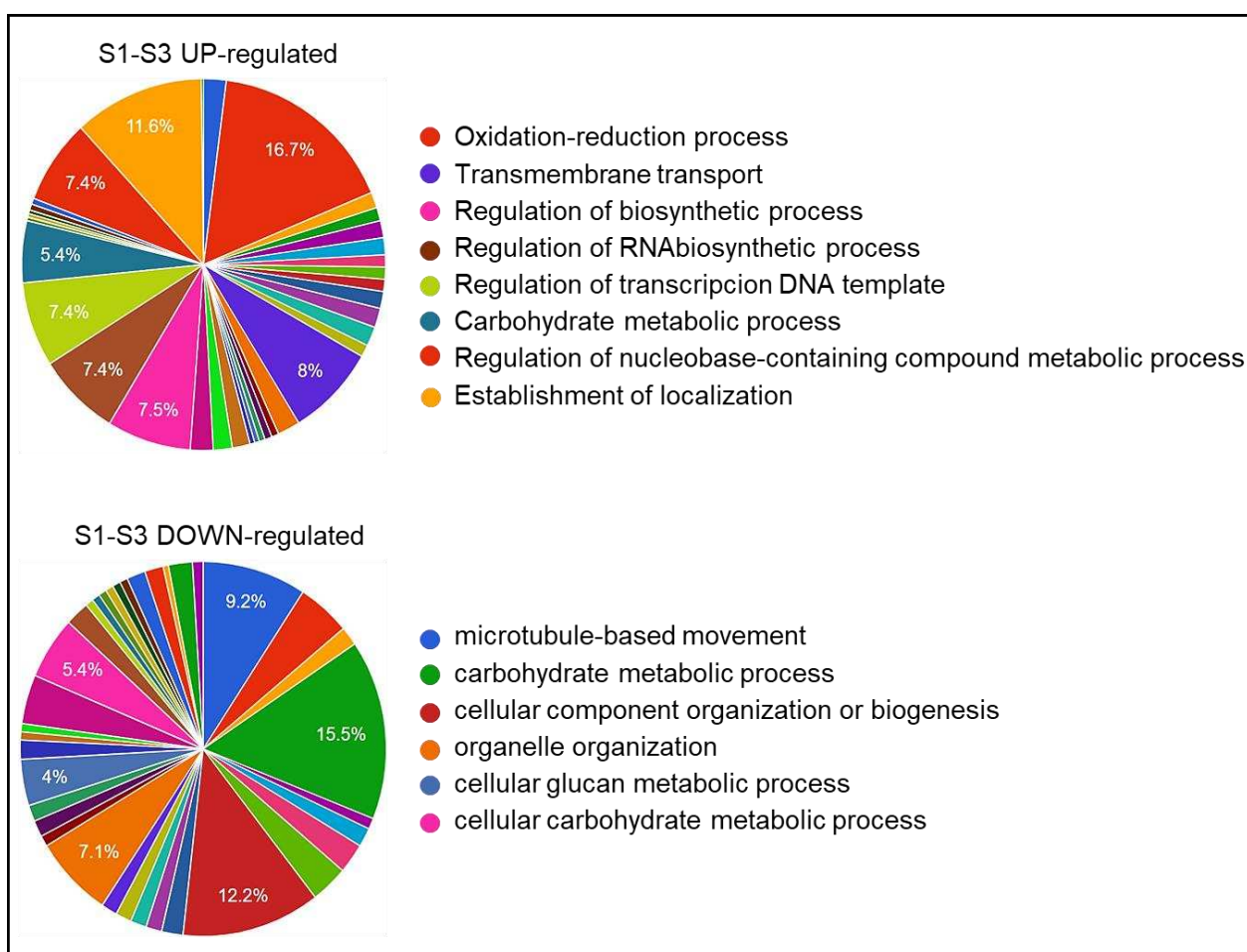


Figura 7: Categorização funcional de genes diferencialmente expressos baseado em processo biológico, a partir do contraste 520-S1 vs 520-S3. A figura ilustra a distribuição de genes diferencialmente expressos entre as categorias funcionais conforme definido em Ontologia de Genes processo biológico. Cada processo biológico é representado por uma cor e um percentual no gráfico. São demonstrados dois gráficos, um para genes diferencialmente expresso *up* regulados e outro para genes *down* regulados e apenas as categorias mais bem representadas são apresentadas ao lado da figura.

Destacadamente, nesta categoria, foi a super representação de genes Cytochrome P450 monooxigenase (CYP) entre os genes *up* regulados (Eucgr.A00523, Eucgr.A01644, Eucgr.A02188, Eucgr.B00199, Eucgr.B00200, Eucgr.B00204, Eucgr.B01014, Eucgr.B01015, Eucgr.B02321, Eucgr.B03132, Eucgr.B03135, Eucgr.B03209, Eucgr.B03265, Eucgr.B03595, Eucgr.B04030, Eucgr.B04037, Eucgr.C00484, Eucgr.C00666, Eucgr.C04098, Eucgr.C04102, Eucgr.D00203, Eucgr.D00209, Eucgr.D01952, Eucgr.D02012, Eucgr.D02168, Eucgr.D02171, Eucgr.E01149, Eucgr.E02618, Eucgr.E03679, Eucgr.E03823, Eucgr.F00146, Eucgr.F00147, Eucgr.F00151, Eucgr.F00153, Eucgr.F00173, Eucgr.F00225, Eucgr.F00492, Eucgr.F01386, Eucgr.F01387, Eucgr.F01492, Eucgr.F01493, Eucgr.F01548, Eucgr.F02072, Eucgr.F02396, Eucgr.F02404, Eucgr.H01953, Eucgr.H01954, Eucgr.H01958, Eucgr.H01963, Eucgr.H04841, Eucgr.H04878, Eucgr.H04879, Eucgr.I00335, Eucgr.I00744, Eucgr.I00749, Eucgr.I00783, Eucgr.I01215, Eucgr.I01217, Eucgr.J00084, Eucgr.J00095, Eucgr.J01402, Eucgr.J02384, Eucgr.K02083, Eucgr.K02211, Eucgr.K02212, Eucgr.K02924, Eucgr.L01403, Eucgr.L02217, Eucgr.L02324, Eucgr.L03497). Estes genes CYP representam aproximadamente 12,23% da superfamília de CYP em eucalipto e tem sido descritos como componentes relevantes do sistema de detoxicação de todos os organismos, sendo que diversos membros são regulados positivamente em diversas condições de estresses fisiológicos (Kim et al., 2013). Estes resultados são consistentes com um estado de estresse associado ao distúrbio fisiológico. CYP450 também está envolvido em diversos processos metabólicos, relacionados à síntese de esteroides e detoxificação de compostos tóxicos, consistente com a regulação positiva das vias de biossíntese de brassinosteróides, diterpenóides e flavonoides, que se observa especificamente em caules suscetíveis sintomáticos (Tabela 3).

Embora não categorizado em função biológica, um conjunto de genes de morte celular foram regulados positivamente (Tabela 4). Neste caso, observa-se que a via de morte celular induzida por múltiplos estresses e mediada pelas proteínas NRPs foi especificamente induzida nos troncos sintomáticos e necróticos com um aumento considerável na expressão do componente a jusante, a enzima VPE (“*vacuolar processing enzyme*”), que age como um efetor de um tipo de morte celular programada específica de plantas mediada pelo colapso do

vacúolo (Nishimura & hatsugai, 2011). Também relevante, foi a indução robusta do homólogo do fator de transcrição ANAC029 (AtNAP), mediador de morte celular programada induzida por senescência foliar natural e em resposta a múltiplos estresses, podendo estar associado ao fenótipo de morte celular dos troncos sintomático (Rogers, 2015). Adicionalmente, como gene alvo controlado por ANAC029, destaca-se SAG113, uma fosfatase 2C, que regula o movimento de estômatos e a perda de água nas folhas, um fenótipo observado em folhas de plantas suscetíveis apresentando os sintomas de distúrbio fisiológico (Zhang & Gan, 2012). Estes resultados indicam que o estabelecimento do estado de distúrbio fisiológico induz o fenômeno de morte celular, culminando em necrose no caule apical e, provavelmente, levando à murcha das folhas, em plantas acometidas pelo distúrbio fisiológico.

Devido ao cultivo em campo aberto, levou-se também em consideração possíveis interações da planta com o meio biótico. Nesta categoria biológica, alguns genes envolvidos no sistema imune inato da planta relacionados com PTI (*PAMP-triggered immunity*), incluindo o receptor de flagelina FLS2, e o co-receptor de PTI, SERK4, e componentes a jusante na via, como CDPK1, MEKK1, WRKY22 e outros foram *up*-regulados em caules 520 sintomáticos (Figura 8, Tabela 5). Estes resultados sugerem que infecções oportunistas de bactérias podem estar acometendo os caules com os sintomas de distúrbio fisiológico. Consistente com esta interpretação, a indução do sistema imune inato foi específica para o contraste 520-S1 vs 520-S3, não tendo sido observado no contraste 1404 R1 vs 1404-S3. Consequentemente, o perfil genético que fundamenta resistência em 1404 não deve estar ligado à imunidade inata de plantas.

Considerando os genes reprimidos (*down* regulados) nos caules sintomáticos (520-S1), predominaram as seguintes categorias biológicas, metabolismo de carboidratos (15.5%), biogêneses e organização de componentes celulares (12.5%), componentes de movimento de microtúbulos (9.2%), organização de organelas (7.1%), metabolismo de polissacarídeos (4.0%), Replicação do DNA e reparo (4.7%) (Figura 7).

Tabela 3. Vias de biossíntese *up*-reguladas especificamente no contraste 520-S1 vs 520-S3

Genes da via de biossíntese de brassinosteróide

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT2G26710	Eucgr.C04102	53.46	2.08	0.00	BAS1
AT2G26710	Eucgr.C04098	25.06	1.92	0.00	BAS1
AT2G26710	Eucgr.L02324	102.88	2.12	0.04	BAS1
AT2G26710	Eucgr.C04098	25.06	1.92	0.00	BAS1
AT3G30180	Eucgr.J01402	6.07	2.68	0.03	BR6OX2
AT4G36380	Eucgr.J02384	283.25	1.40	0.00	CYP90D1

Genes da via de biossíntese de diterpenóides

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT1G02400	Eucgr.K03561	71.86	2.37	0.00	GA2OX6
AT2G32440	Eucgr.K02924	110.01	2.30	0.00	CYP88A3
AT2G32440	Eucgr.D02168	174.45	3.01	0.00	CYP88A3
AT2G32440	Eucgr.D02171	52.91	3.17	0.00	CYP88A3
AT5G25900	Eucgr.F02396	47.78	1.34	0.03	GA3

Genes da via de biossíntese de flavonoides

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT2G40890	Eucgr.A02188	1018.30	1.73	0.01	CYP98A3
AT5G07990	Eucgr.I00335	266.76	2.66	0.00	TT7
AT5G07990	Eucgr.K02211	162.37	3.36	0.00	TT7
AT5G13930	Eucgr.D01635	927.34	4.65	0.00	TT4

Tabela 4. Genes envolvidos em morte celular programada

ID	log2foldchange	Pvalor ajustado	Descrição
Eucgr.E02876	3.21	0.000740459030104	AT1G17020.1; ATSRG1,SRG1
Eucgr.I00575	1.23	2.77231413391e-05	AT1G17020.1; ATSRG1,SRG1
Eucgr.E02899	2.13	0.0295366968439	AT1G17020.1; ATSRG1,SRG1
Eucgr.I00101	2.50	0.029717655	AT1G01720.1; ANAC002,ATAF1; NAC
Eucgr.B03208	13.46	3.26345765586e-16	AT1G69490.1; ANAC029,ATNAP,NAP; NAC-like, activated by AP3/PI
ucgr.A02070	2.26	5.47828723948e-06	AT3G04070.1; anac047,NAC047; NAC domain containing protein 47
Eucgr.B00529	2.84	5.23475584447e-09	AT5G61430.1; ANAC100,ATNAC5,NAC100; NAC domain containing protein 100
Eucgr.B03537	1.40	0.0149575511958	AT5G13180.1; ANAC083,NAC083,VNI2; NAC domain containing protein 83
Eucgr.D00665	7.20	1.67418011205e-10	AT1G61110.1; anac025,NAC025; NAC domain containing protein 25
Eucgr.F01091	2.79	1.96476669734e-05	AT3G15510.1; ANAC056,ATNAC2,NAC2,NARS1; NAC domain containing protein 2
Eucgr.F01449	2.05	0.0474140379725	AT1G56010.2; anac021,ANAC022,NAC1; NAC domain containing protein 1
Eucgr.F04341	2.26	0.0011360926289	AT1G01720.1; ANAC002,ATAF1; NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
Eucgr.G02486	3.69	2.23462350998e-08	AT1G69490.1; ANAC029,ATNAP,NAP; NAC-like, activated by AP3/PI
Eucgr.H03387	1.60	0.0153873314179	AT5G13180.1; ANAC083,NAC083,VNI2; NAC domain containing protein 83
Eucgr.I02695	4.55	0.000481283704482	AT5G64530.1; ANAC104,XND1; xylem NAC domain 1
Eucgr.K01228	2.63	0.000128129827113	AT2G43000.1; anac042,NAC042; NAC domain containing protein 42
Eucgr.J03189	2.43	0.0307120188604	AT2G25940.1; ALPHA-VPE,ALPHAVPE; alpha-vacuolar processing enzyme
Eucgr.B03492	1.84	9.25665064566e-06	PTHR12298 (PCDC2 PROGRAMMED CELL DEATH PROTEIN 2 -RELATED)
Eucgr.H04638	2.03	0.000434880968671	PTHR12298 (PCDC2 PROGRAMMED CELL DEATH PROTEIN 2 -RELATED)
Eucgr.K01055	1.52	0.0196129786515	PTHR12626 (PROGRAMMED CELL DEATH 4),
Eucgr.K02766	1.26	3.53919311225e-06	PTHR12626 (PROGRAMMED CELL DEATH 4),
Eucgr.F03472	3.04	0.0321352755537	PF04520 (Senescence regulator)
Eucgr.I01873	2.48	0.00987047893745	AT5G14930.2; SAG101; senescence-associated gene 101
Eucgr.K03248	4.52	0.000234211367753	PF04520 (Senescence regulator)
Eucgr.L02488	3.33	0.0204290874677	PF04520 (Senescence regulator)
Eucgr.E01896	2.77	0.000217425310007	PTHR10044 (INHIBITOR OF APOPTOSIS),
Eucgr.E01902	2.61	7.67617250547e-05	PTHR10044 (INHIBITOR OF APOPTOSIS),
Eucgr.E01904	2.47	0.00135323218813	PTHR10044 (INHIBITOR OF APOPTOSIS)
Eucgr.I02659	1.94	0.00700566734938	PF00656 (Caspase domain) metacaspase 1

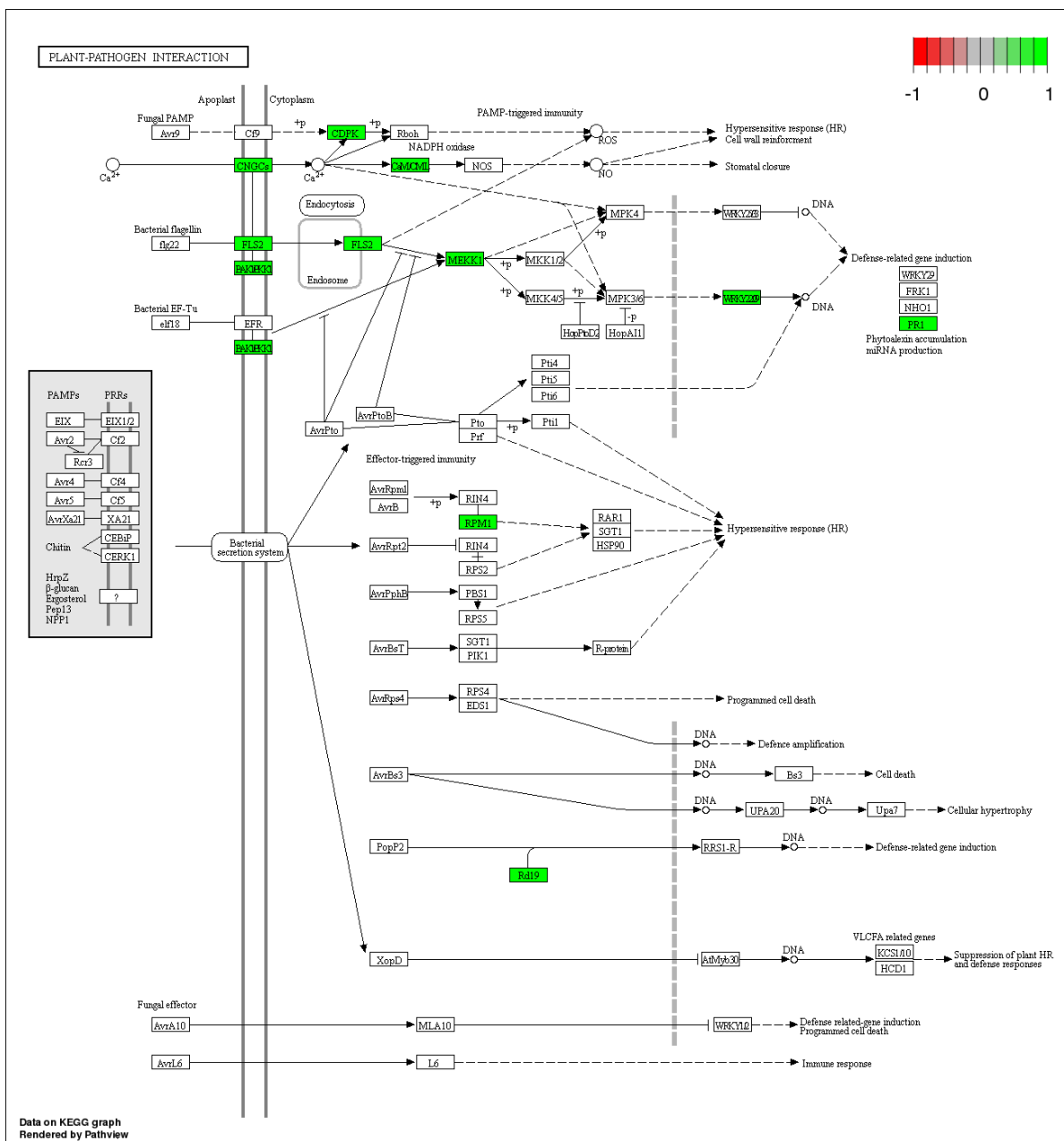


Figura 8: Genes envolvidos em PTI são induzidos pelo distúrbio fisiológico em caules sintomáticos do clone 520 (amostra 520-S1). Genes em vermelho são reprimidos e em verde induzidos, utilizando-se os critérios de \log_2 fold change = 1 e p-valor de 0.05.

Tabela 5. Indução de genes envolvidos em PTI no contraste 520-S1 vs 520-S3.

AIR10	ID	basemean	log2foldchange	p-valor	Nomes
AT1G01340	Eucgr.E00062	34.02	1.58	0.01	CNGC10
AT1G01340	Eucgr.E00059	25.36	1.74	0.01	CNGC10
AT1G18210	Eucgr.F00233	387.06	1.92	0.05	AT1G05990
AT2G14610	Eucgr.L01707	445.84	2.81	0.04	AT1G50060
AT2G14610	Eucgr.G01137	399.75	2.92	0.03	AT1G50060
AT3G07040	Eucgr.D02631	21.59	1.45	0.02	RPM1
AT4G01250	Eucgr.D02324	334.19	2.56	0.00	WRKY22
AT4G01250	Eucgr.E00098	29.08	3.58	0.02	WRKY22
AT4G08500	Eucgr.G00627	139.05	1.70	0.00	MEKK1
AT4G33430	Eucgr.K01601	36.21	3.07	0.00	SERK4
AT4G39090	Eucgr.F01612	5841.68	1.49	0.00	AT2G21430
AT5G46330	Eucgr.D00318	119.00	4.85	0.00	FLS2
AT5G66210	Eucgr.I02347	450.18	1.29	0.00	CDPK1

Entre os genes regulados negativamente, na categoria biogênese, destaca-se uma repressão massiva de genes envolvidos na síntese de ácidos graxos, incluindo ácidos graxos de cadeia longa e ceras (Figura 9, Tabela 6). Estes genes amplamente reprimidos incluem *FATB* que codifica FATTY ACYL-ACP THIOESTERASES B, *ACCI*, que codifica ACETYL-COA CARBOXYLASE, *LACS2*, LATERAL ROOT DEVELOPMENT 2, também designado LONG-CHAIN ACYL-COA SYNTHETASE 2, *FABI*, FATTY ACID BIOSYNTHESIS 1, *MOD1*, MOSAIC DEATH 1, também designado ENOYL-ACP REDUCTASE 1, *EMB3147*, EMBRYO DEFECTIVE 3147, também designado, MALONYL COA-ACP MALONYLTRANSFERASE, *FATA1* que codifica FatA acyl-ACP thioesterase, e *FTMI*, FLORAL TRANSITION AT THE MERISTEM1. Estes genes, essenciais para síntese de ácidos graxos de cadeia longa, estão envolvidos em diversas etapas

de desenvolvimento e crescimento de plantas, como síntese de ceras e cutina nos caules e folhas, formação da raiz lateral, controle de divisão celular na região de meristemas (Bonaventure et al.,2003). Reduções ou bloqueios na síntese de ácidos graxos saturados têm sido associados a declínios na taxa de crescimento de plantas, na síntese de ceras e cutina, um dos principais componentes da cutícula da planta e indução de morte celular. Em plantas silenciadas para o gene *FATB*, o carregamento total de ceras na folha é reduzido 20% e, nos caules, a redução chega a 50%, o que implica *FATB* no suprimento de ácidos graxos saturados para biossíntese de ceras (Bonaventure et al.,2003). Mutantes em *ACCI* exibem divisões celulares descoordenadas, uma organização aberrante da região apical do embrião e desenvolvimento anormal de caules e raízes, sendo deficientes à tolerância ao congelamento após aclimação ao frio (Baud et al., 2003). O gene *LACS2* catalisa a síntese de omega-hydroxy fatty acyl-CoA, intermediário na via de síntese de cutina, enquanto que *MOD1* codifica enoyl-ACP reductase, um componente do complex sintase de ácidos graxos (Schnurr & Shockey, 2004). Redução na expressão de *MOD1* causa morte celular prematura, um fenótipo característico de caules 520-S1. Além disso, plantas mutantes para *MOD1* possuem nível de lipídeo reduzido e defeitos morfológicos pleiotrópicos incluindo folhas cloróticas e com formato anormal, que mimetizam o fenótipo de folhas nas amostras 520-S4.

Embora todas as mutações nos genes da biossíntese de ácidos graxos tenham sido geradas e analisadas em *Arabidopsis*, o fenótipo de plantas de eucalipto exibindo distúrbio fisiológico pode ser amplamente explicado pela redução na expressão de genes essenciais na biossíntese de ácidos graxos de cadeia longa.

Tabela 6. Repressão de genes envolvidos na síntese de ácidos graxos no contraste 520-S1 vs 520-S3

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT1G08510	Eucgr.C01082	377.56	-2.30	0.00	FATB
AT1G08510	Eucgr.C04139	2356.73	-1.20	0.00	FATB
AT1G24360	Eucgr.H04666	130.54	-3.04	0.00	AT1G24360
AT1G36160	Eucgr.F00484	3036.56	-1.25	0.02	ACC1
AT1G49430	Eucgr.G00490	665.95	-3.14	0.00	LACS2
AT1G74960	Eucgr.C00778	194.52	-1.39	0.02	FAB1
AT2G05990	Eucgr.A02448	453.80	-1.04	0.01	MOD1
AT2G30200	Eucgr.J01859	363.50	-1.03	0.02	EMB3147
AT2G43710	Eucgr.H00296	5.89	-5.89	0.02	FTM1
AT3G25110	Eucgr.H00774	952.19	-2.68	0.00	FaTA

Na categoria Replicação do DNA e Reparo, predomina a regulação negativa de genes relacionados ao sistema de reparo do DNA, indicando que falhas no sistema pode ter sido o estímulo para ativação de vias de morte celular nos caules sintomáticos (520-S1) (Figura 10, Tabela 7). As plantas são particularmente vulneráveis a fatores que danificam DNA. Desta forma, elas evoluíram uma rede complexa de mecanismos para detecção de danos no DNA e reparos, dedicados a assegurar a estabilidade genômica por meio de remoção das lesões no DNA e reconstituição da informação genética original (Kireva et al., 2013). Se o sistema de reparo de DNA é comprometido, então instabilidade genômica gera um sinal para ativar vias de morte celular. O aparato de reparo celular, conta com a presença de proteínas que após reconhecimento do sítio danificado, desencadeiam a interrupção do ciclo celular, enquanto enzimas relacionadas diretamente com a correção do dano são recrutadas (Kireva et al., 2013). Este cenário é consistente com as variações de expressão gênica observadas nos caules sintomáticos pelo estabelecimento do estado de distúrbio fisiológico.

Os genes diferencialmente expressos na comparação entre 520-S2 e 520-S3 deve representar o transcriptoma induzido pela recuperação de sintomas, uma vez que o caule que emerge pela perda da dominância apical (amostra S2) em substituição ao caule necrosado (amostra S1) não apresenta sintomas do distúrbio fisiológico. A categorização funcional para os genes diferencialmente expressos, demonstram uma predominância das seguintes categorias biológicas para os genes *up* regulados: genes relacionados a processo de oxidação-redução (29.8%), transcrição (16.3%), biossíntese de RNA (16.3%). Em relação aos genes *down* regulados, denota um percentual maior para componentes de organização celular (15.4%), replicação do DNA (5.7%), ciclo celular (7.3%) e metabolismo de carboidrato (6.5%) (Figura 11). O estado de transição dessas amostras S2 é evidenciado ao se comparar expressão diferencial de genes específicos do contraste 520-S1 vs 520-S3 com 520-S2 vs 520-S3. Observa-se, por exemplo, que a repressão massiva dos genes da síntese de ácidos graxos em 520-S1 já não é observada em S2. Assim também, o sistema de reparo de DNA já não apresenta repressão acentuada de genes em S2.

A análise de categorização funcional dos genes DE para o contraste 1404-R1 vs 520-S3, por meio de Gene Ontology (GO) revelou que os genes regulados positivamente, apresentam como categorias biológicas mais enriquecidas, biossíntese celular (20.4%),

biossíntese de macromoléculas (16.6%) e tradução (7.7%), enquanto os genes regulados negativamente estão super representados nos processos biológicos com os termos processos de modificação de proteínas (10.3%), regulação biológica (7.1%), processos metabólicos primários (20.0%) e metabolismo de proteína celular (11.2%) (Figura 12). Em destaque, está a predominância de genes de proteínas ribossomais e da maquinaria de tradução no transcriptoma *up* regulado das amostras R1 (Figura 13). Esta parece ser uma característica comparativa entre genótipos (clones 520 e 1404), porque não se observa expressão diferencial desses genes da maquinaria de tradução pelo distúrbio fisiológico (contraste 520-S1 vs 520-S3). Também significativo foi a *up*-regulação de genes envolvidos no reparo do DNA nas amostras 1404-R1 em comparação a 520-S3 (Figura 14). Em contraste à expressão diferencial de genes de proteínas ribossomais, uma melhor eficiência de reparo de DNA pode ser atribuída, pelo menos em parte, ao mecanismo de resistência do clone 1404, uma vez que o estado de distúrbio fisiológico promoveu repressão de genes de reparo do DNA (Figura 10).

Nos genes regulados negativamente em 1404-R1, destacam-se genes envolvidos no sistema imune de plantas, tanto em PTI quanto em ETI (*effector triggered immunity*) (Figura 15). A Tabela 8 descreve a regulação diferencial de genes de PTI e ETI em caules do clone 1404 (amostra R1), já caracterizados como funcionais do sistema imune de plantas. Os genes de resistência (genes R) *down* regulados que ativam ETI em reconhecimento específico com efetores de bactérias (genes *avr*) já caracterizados incluem RIN4, RPM1, RPS5, SGT1 e EDS1 (Figura 15). Além desses, foi observado repressão extensiva na expressão de genes que codificam proteínas de resistências com o domínio TIR (*toll-like domain*) na configuração estrutural Toll-NBS-LRR (NBS, *nucleotide binding site* e LRR, *leucine-rich repeats*). Estes resultados indicam que o mecanismo de resistência ao distúrbio fisiológico não pode ser atribuído a resistência basal a patógenos gerais ou específicos e, portanto, o estímulo ambiental que induz o estado de distúrbio fisiológico não corresponde a estress biótico. Além disso, em termos de metagenômica funcional, não se identificou uma classe de patógenos predominante que poderia estar causando o distúrbio fisiológico nas amostras analisadas por RNA-Seq (dados não mostrados).

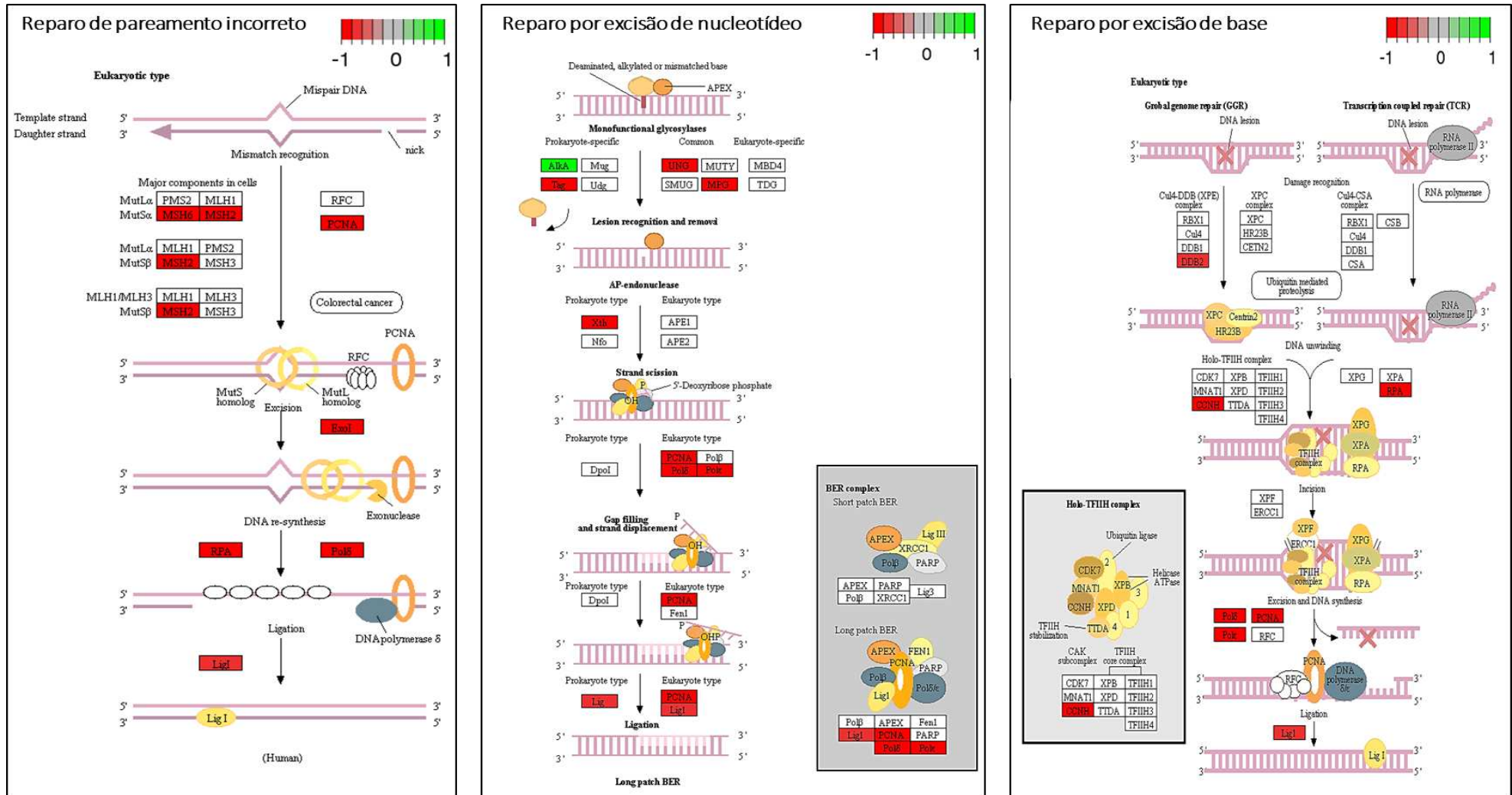


Figura 10 : Genes envolvidos no reparo de DNA diferencialmente expressos em 520-S1 comparado com 520-S3. Genes em vermelho são reprimidos e em verde induzidos, utilizando-se os critérios de \log_2 fold change = 1 e p-valor de 0.05

Tabela 7. Genes envolvidos no reparo de DNA diferencialmente expressos em 520-S1 comparado com 520-S3

03430 Mismatch repair

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT1G08260	Eucgr.B03882	81.33	-2.90	0.00	TIL1
AT1G13635	Eucgr.B03112	178.49	-6.07	0.00	AT1G13635
AT2G41460	Eucgr.J01377	14.40	-2.06	0.04	ARP
AT3G12040	Eucgr.K01848	49.67	-1.07	0.02	AT3G12040
AT3G18630	Eucgr.J00843	176.61	-1.42	0.05	UNG
AT5G63960	Eucgr.I00173	97.14	-1.01	0.01	POLD4

03420 Nucleotide excision repair

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT1G08260	Eucgr.B03882	81.33	-2.90	0.00	TIL1
AT2G24490	Eucgr.C02095	34.23	-2.25	0.00	AT1G03810
AT5G27620	Eucgr.H01039	44.90	-1.03	0.01	CYCH;1
AT5G63960	Eucgr.I00173	97.14	-1.01	0.01	POLD4

03410 Base excision repair

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT1G08260	Eucgr.B03882	81.33	-2.90	0.00	TIL1
AT1G13635	Eucgr.B03112	178.49	-6.07	0.00	AT1G13635
AT2G41460	Eucgr.J01377	14.40	-2.06	0.04	ARP
AT3G12040	Eucgr.K01848	49.67	-1.07	0.02	AT3G12040
AT3G18630	Eucgr.J00843	176.61	-1.42	0.05	UNG
AT5G63960	Eucgr.I00173	97.14	-1.01	0.01	POLD4

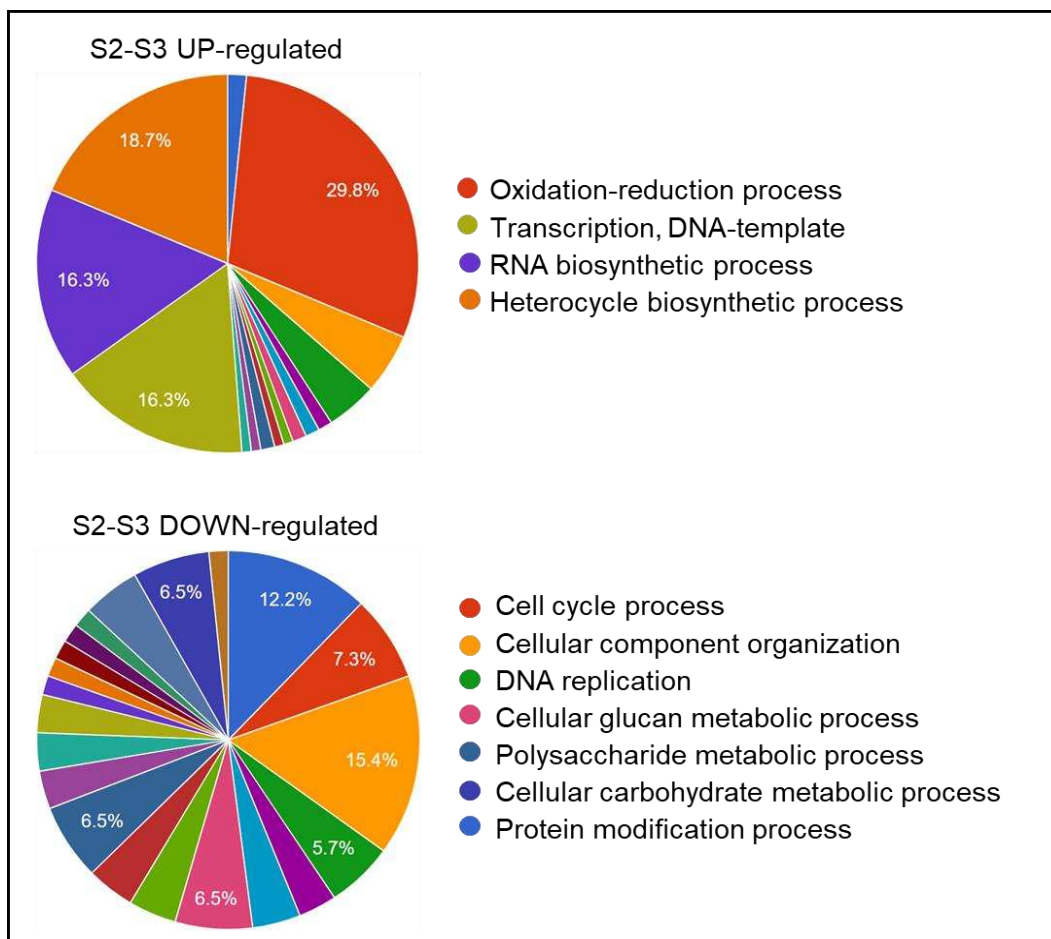


Figura 11: Categorização funcional de genes diferencialmente expressos baseado em processo biológico, a partir do contraste 520-S2 vs 520-S3. A figura ilustra a distribuição de genes diferencialmente expressos entre as categorias funcionais conforme definido em Ontologia de Genes-processo biológico. Cada processo biológico é representado por uma cor e um percentual no gráfico. São demonstrados dois graficos, um para genes diferencialmente expresso *up* regulados e outro para genes *down* regulados.

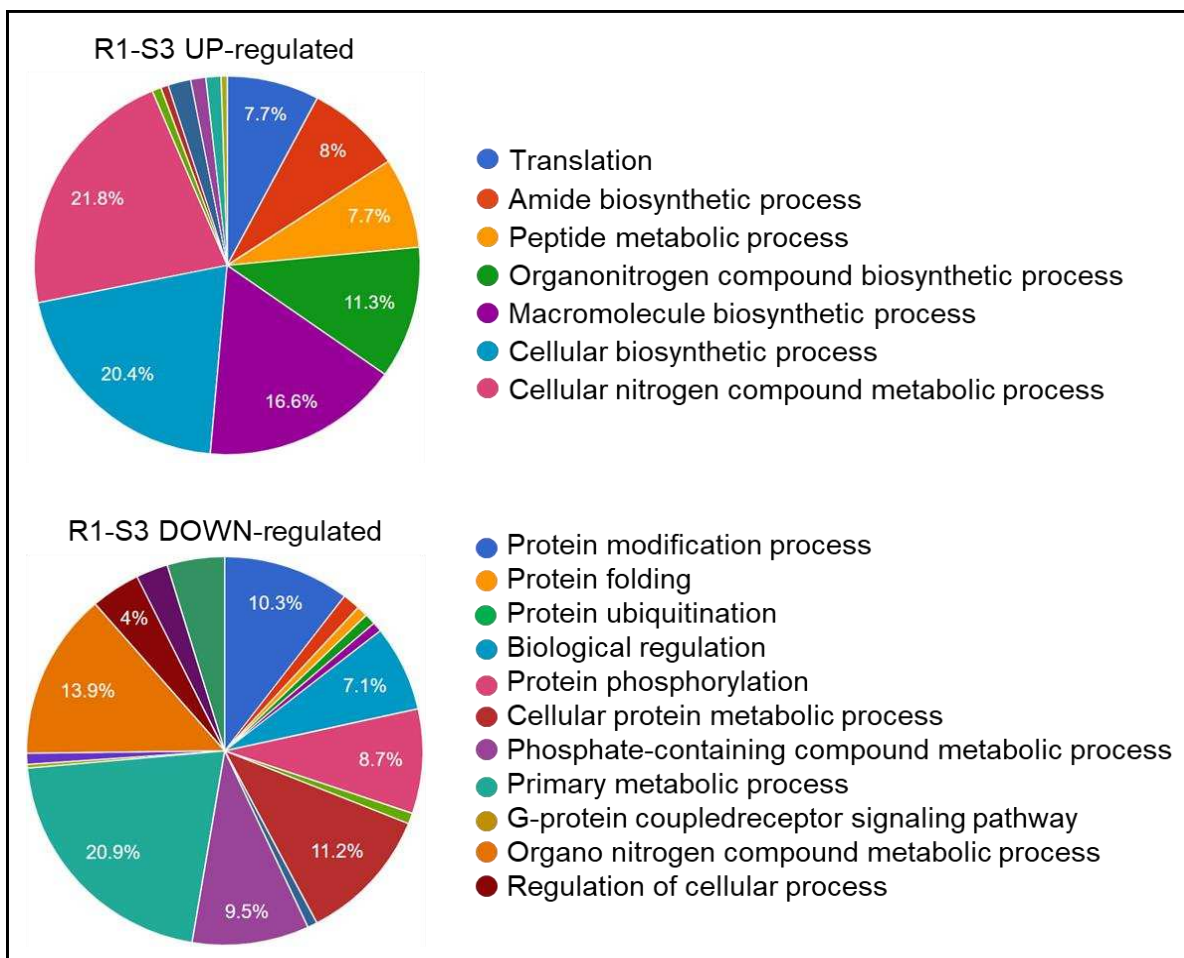


Figura 12: Categorização funcional de genes diferencialmente expressos baseado em processo biológico, a partir do contraste 1404-R1 vs 520-S3. A figura ilustra a distribuição de genes diferencialmente expressos entre as categorias funcionais conforme definido em Ontologia de Genes-processo biológico. Cada processo biológico é representado por uma cor e um percentual no gráfico. São demonstrados dois gráficos, um para genes diferencialmente expresso *up* regulados e outro para genes *down* regulados.

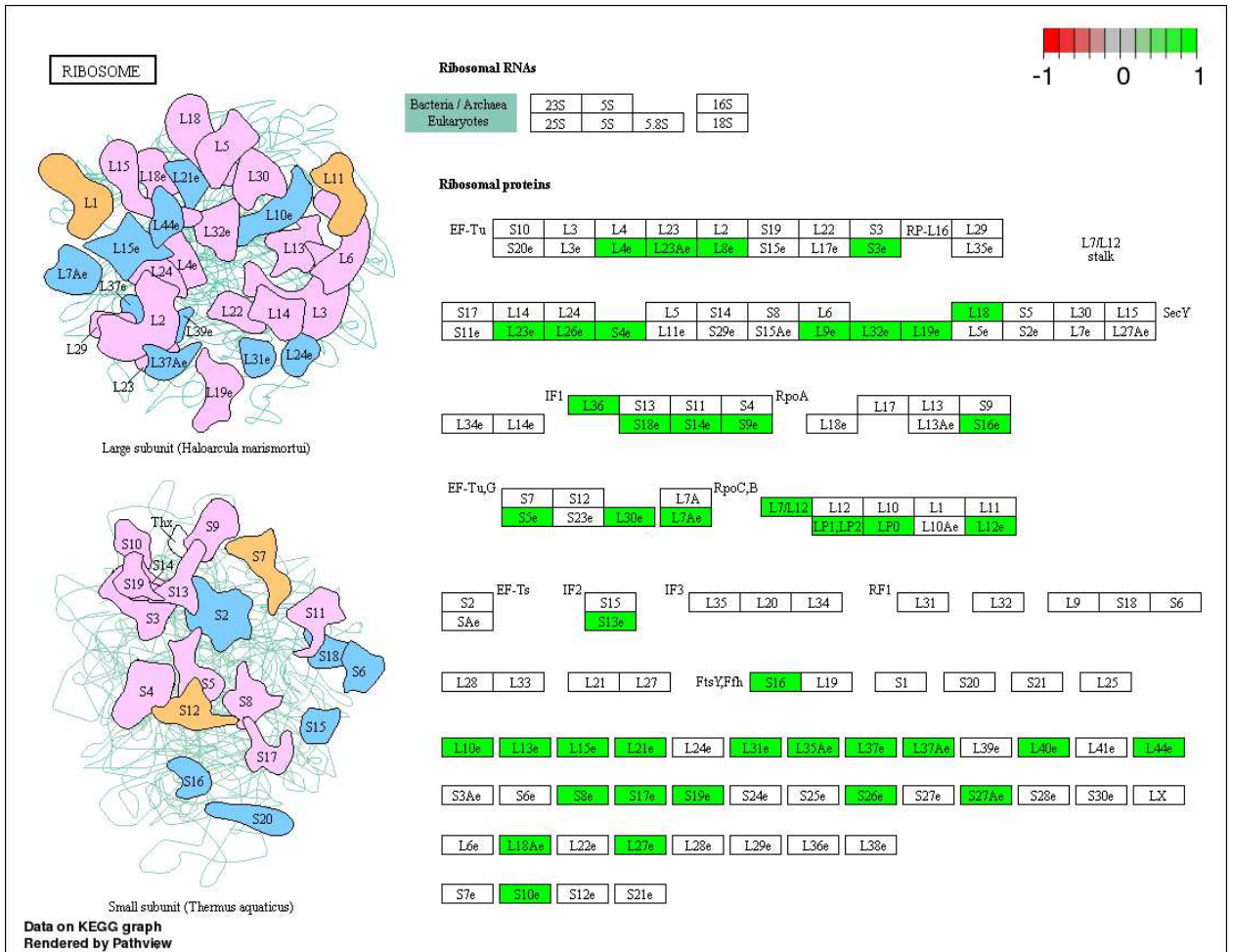


Figura 13: Genes codificam proteínas ribossomais diferencialmente expressos em 1404-R1 comparados com 520-S3. Genes em vermelho são reprimidos e em verde induzidos, utilizando-se os critérios de \log_2 -fold-change = 1, e p-valor de 0,05.

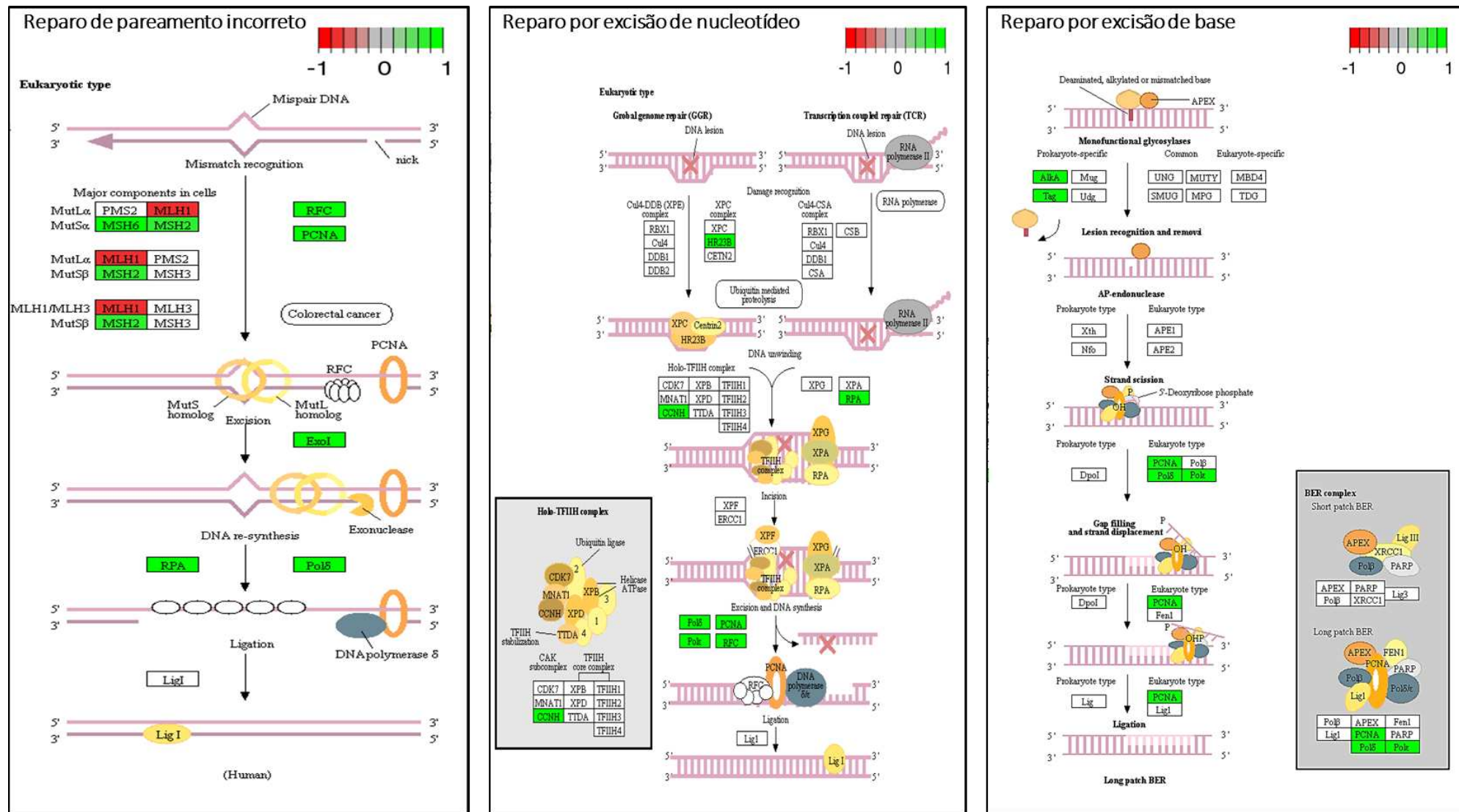


Figura 14: Genes envolvidos no reparo de DNA diferencialmente expressos em 1404-R1 comparado com 520-S3. Genes em vermelho são reprimidos e em verde induzidos, utilizando-se os critérios de \log_2 fold change = 1 e p-valor de 0.05

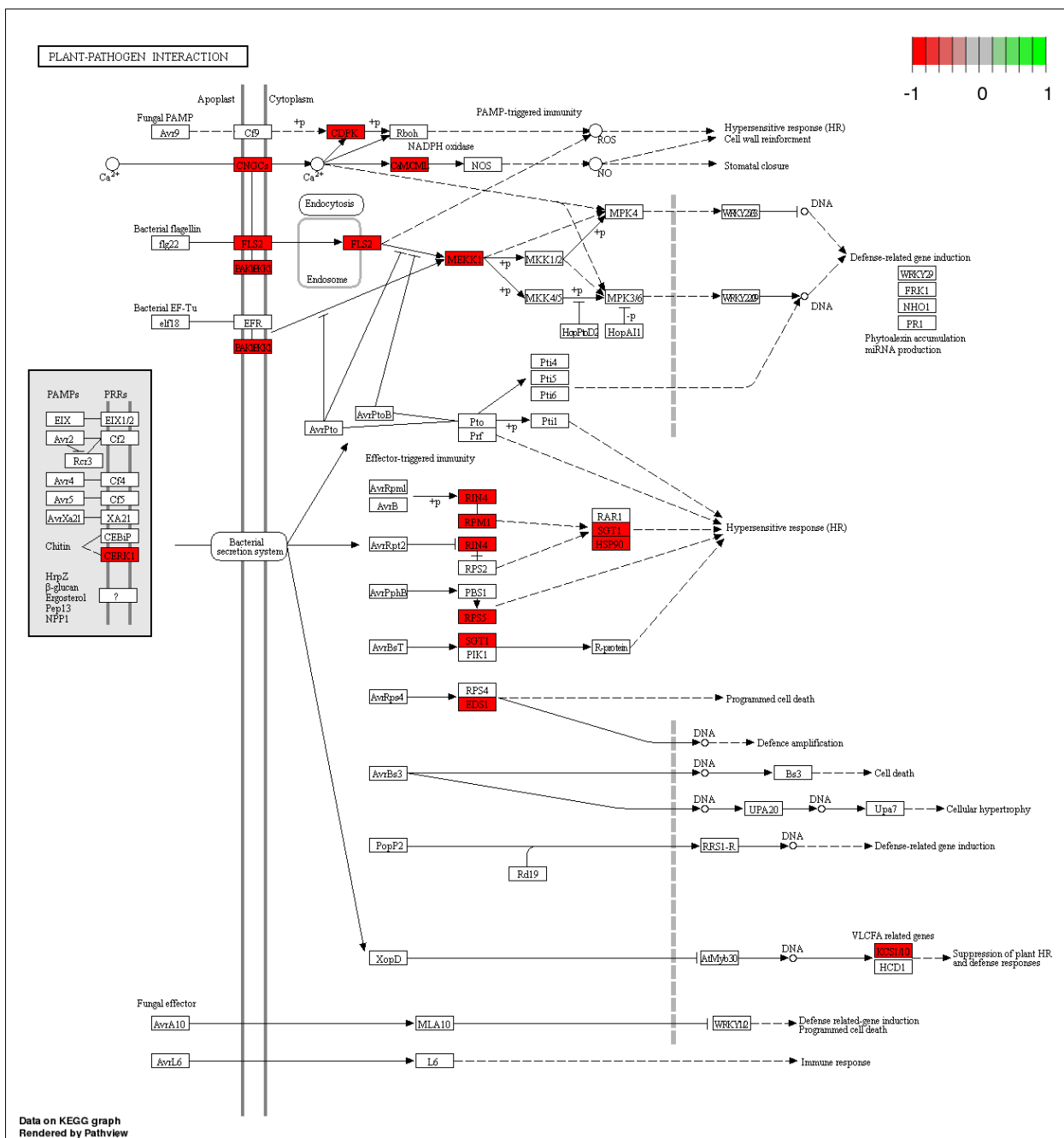


Figura 15: Repressão extensiva de genes do sistema auto-imune inato de plantas (PTI e ETI) em 1404-R1 comparado com 520-S3. Genes em vermelho são reprimidos e em verde induzidos, utilizando-se os critérios de \log_2 -fold-change = 1, e p-valor de 0,05.

Tabela 8. Genes envolvidos em PTI e ETI reprimidos em caules do clone 1404 (R1)

TAIR10	ID	basemean	log2foldchange	p-valor	labels
AT1G12220	Eucgr.F01108	26.41	-1.96	0.02	RPS5
AT1G19440	Eucgr.F02683	15.19	-5.18	0.04	KCS1
AT1G21550	Eucgr.E01001	85.37	-3.89	0.00	AT1G05990
AT1G35670	Eucgr.F00807	935.32	-1.16	0.04	CDPK1
AT2G04030	Eucgr.B03110	2434.18	-1.59	0.00	CR88
AT2G13790	Eucgr.K01609	112.67	-2.32	0.00	SERK4
AT3G07040	Eucgr.G01440	71.48	-4.30	0.00	RPM1
AT3G21630	Eucgr.A01103	734.46	-1.45	0.00	CERK1
AT3G25070	Eucgr.H00779	920.57	-3.57	0.00	RIN4
AT3G48090	Eucgr.K02886	1005.06	-1.48	0.04	AT3G48080
AT3G48090	Eucgr.L03388	162.25	-1.66	0.03	AT3G48080
AT4G08500	Eucgr.E03542	87.73	-5.12	0.00	MEKK1
AT4G08500	Eucgr.F03743	1631.52	-1.49	0.00	MEKK1
AT4G11260	Eucgr.D01807	6606.12	-1.53	0.00	SGT1B
AT4G30360	Eucgr.C02438	1162.75	-1.11	0.02	CNGC10
AT5G46330	Eucgr.C03760	16.50	-4.86	0.00	FLS2

Com a finalidade de delinear genes diferencialmente expressos em R1 que pudessem contribuir com o efeito resistência, foi utilizado um critério de busca que identificasse genes diferencialmente regulados (*up* e *down*) em oposição entre os contrastes 520-S1 vs 520-S3 e 1404-R1 vs 520-S3. A premissa seria que genes que conferissem resistência ao distúrbio fisiológico deveriam exibir regulação oposta entre as amostras suscetíveis acometidas pelo distúrbio e as amostras resistentes sem sintomas. Para isso, foi construído diagramas de Venn, a partir dos genes diferencialmente expressos identificados pela comparação dos contrastes 520-S1 vs 520-S3 x 1404-R1 vs 520-S3, utilizando como critérios, $\log_2\text{-fold-change} = 1$, e p-valor de 0,01 e comparando genes *up* regulados no contraste 1404-R1 vs 520-S3 com genes *down* regulados no contraste 520-S1 vs 520-S3

(Figura 16) e vice-versa (Figura 17). Em um total de de 3070 genes diferencialmente expressos em direções opostas, 60 genes foram identificados como sendo reprimidos em 520-S1 e induzidos em 1404-R1 em comparação com S3 (Figura 16). Consistente com a sintomatologia do distúrbio fisiológico, destaca-se o gene Eucgr.F03178, homólogo de AT3G12920.1, um inibidor de apoptose, além de genes envolvidos na regulação de meristemas, reparo do DNA, e em tolerância a estresses abióticos como genes induzidos por déficit hídrico e calor.

Conforme os resultados da classificação feita por enriquecimento dos genes DE em relação aos processos biológicos, genes envolvidos em processos como morte celular (4,1%), desenvolvimento de meristemas (4,1%), biossíntese de lipídeo (6,2%), regulação negativa da recombinação do DNA (1,0%), processos metabólicos secundários (5,2%), elementos repetidos de DNA (1,0%), desenvolvimento (11,2%), regulação do crescimento (4,1%), regulação do crescimento do meristema (4,1%), ou seja, envolvidos em eventos de desenvolvimento, processos de biossíntese e metabolismo celular, além de reparo do DNA, foram reprimidos nas amostras sintomáticas-S1 e induzidos nas amostras resistentes -R1 (Figura 16). Em contraposição genes envolvidos na resposta a estímulos abióticos (9,8%), regulação positiva de processo celular (4,9%), resposta ao frio (4,9%), regulação de processos biológicos (14,8%), regulação de processos metabólicos de macromoléculas (4,9%), regulação positiva do processo biosintético de RNA (4,9%), regulação positiva da transcrição (4,9%) e resposta a substâncias químicas (11,5%), enquadrados de forma geral em processos biológicos, celular, de regulação e de resposta a estímulos, foram *up* regulados nas amostras sintomáticas-S1 e *down* regulados nas amostras resistentes -R1 (Figura 17). Coletivamente, estes resultados sugerem que o distúrbio fisiológico no clones 520, genótipo suscetível, está associado a alteração negativa no sistema de reparo do DNA, o que provavelmente ativa eventos de morte celular programada no clone 520 , desencadeando processos que envolvem principalmente resposta ao estresse abiótico, síntese de alcaloide, flavonoides e em contrapartida, diminuição de processos relacionados com desenvolvimento, replicação do DNA e metabolismo. Os resultados gerais de variação global de expressão gênica também eliminam a possibilidade de que o distúrbio fisiológico seja ativado por estímulos bióticos.

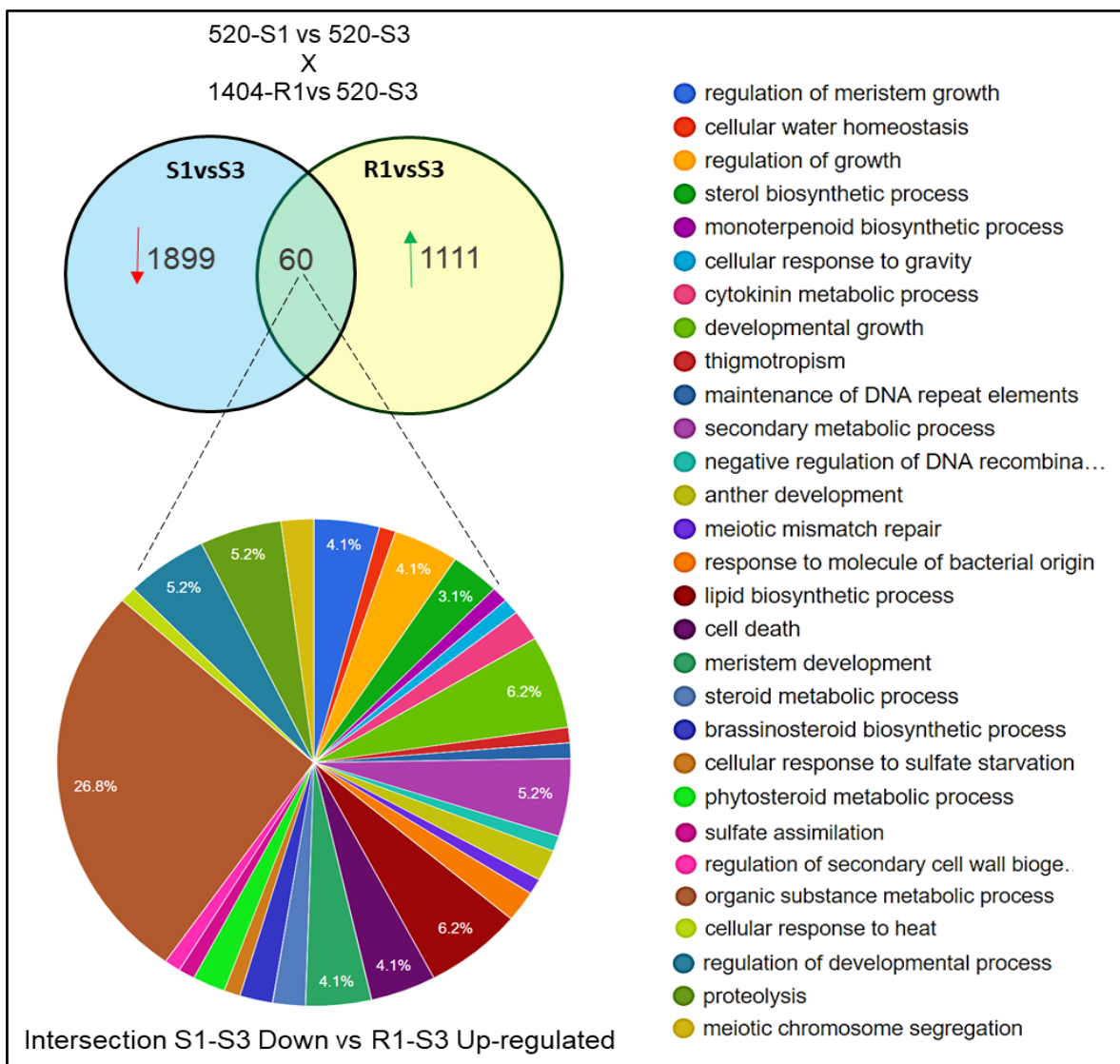


Figura 16: Diagramas de Venn mostrando as relações entre os diferentes conjuntos de genes diferencialmente expressos nos clones 520-S1, 520-S3 e 1404-R1. Regiões em azul claro representam o contraste entre os genes DE *down* regulados em S1 vs S3, regiões em amarelo claro, contraste entre os genes DE *up* regulados R1 vs S3 e a região comum a estas duas áreas, representam os genes de interseção. Pela análise de enriquecimento dos genes DE, em processo biológico, é fornecido gráfico indicando o percentual de cada gene correspondente. Foi utilizado \log_2 -fold-change = 1, e p-valor de 0,01.

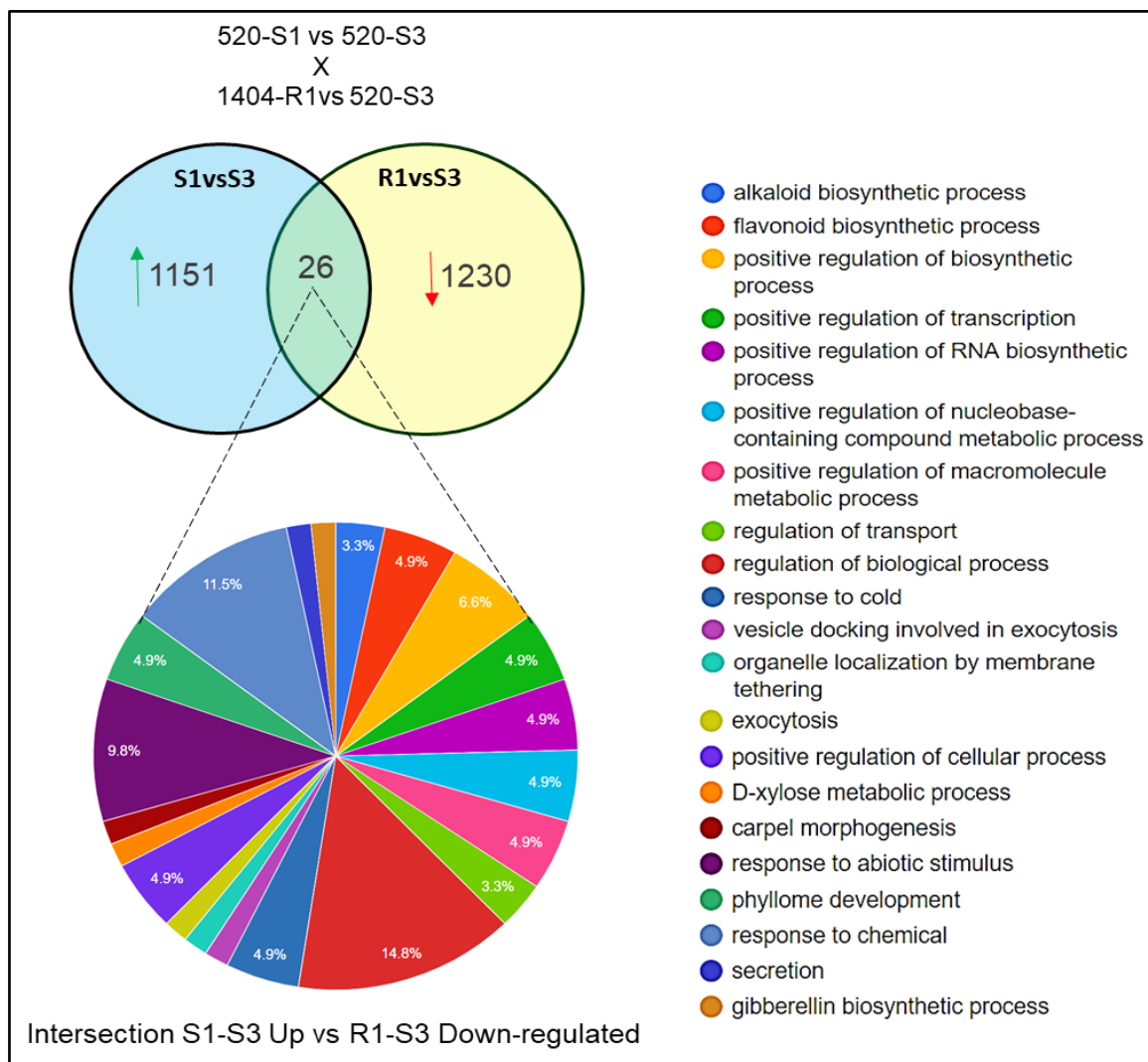


Figura 17: Diagramas de Venn mostrando as relações entre os diferentes conjuntos de genes diferencialmente expressos nos clones 520-S1, 520-S3 e 1404-R1. Regiões em azul claro representam o contraste entre os genes DE *up* regulados em S1 vs S3, regiões em amarelo claro, contraste entre os genes DE *down* regulados R1 vs S3 e a região comum a estas duas áreas, representam os genes de interseção. Por análise de enriquecimento dos genes DE, em processos biológicos, é fornecido gráfico indicando o percentual de cada gene correspondente. Foi utilizado \log_2 -fold-change = 1, e p-valor de 0,01.

3.4 Análise de expressão gênica para os genes candidatos

Para validação dos resultados de RNA-Seq, confirmando a expressão diferencial de um subconjunto de genes responsivos ao estresse foi realizado análises de qRT-PCR das amostras 520-S1, 520-S3, 1404-R1. Foram selecionados três genes da via da síntese de ácidos graxos, três genes envolvidos no sistema de reparo do DNA, além de alguns genes diferencialmente expressos em S1 ou R1, escolhidos aleatoriamente (Tabela 9). Como controle endógeno foram testados os genes *eEF2* (*Eukaryotic translation elongation factor 2*), *ACT* (*Actin-related protein 3*) e *30S* (*30S ribosomal protein S5, chloroplastic*).

A especificidade e eficiência dos *primers* específicos do gene, foi determinada por eletroforese em gel de agarose (RT-PCR) (Figura 18B) e pela análise da curva de *melting*. Cada uma das curvas de *melting* desses genes evidenciaram um único pico em qRT-PCR e as curvas de amplificação das nove repetições (três repetições biológicas x três repetições técnicas) para cada gene apresentou excelente repetibilidade. Esses resultados indicam alta especificidade *dos primers* para seus respectivos genes-alvo. O cálculo da eficiência de amplificação da PCR foi feito para os três genes endógenos, e variaram de 0,9 a 1,18, atendendo aos requisitos para análises com PCR quantitativo (Figura 18C).

De uma maneira geral, todos os genes selecionados apresentaram acúmulos de transcritos consistentes com os resultados do RNA-Seq (Figura 19). O nível de transcritos dos genes envolvidos na síntese de ácido graxo, KASIII, ACC1 e KCS1, foi muito inferior em S1 quando comparados com S3, confirmando a repressão global da síntese de ácidos graxos em caules sintomáticos do clone 520. Não se observaram variações significativas desses transcritos em 1404-R1, confirmando os resultados do RNA-Seq. Quanto aos genes envolvidos no reparo de DNA, DNA Glycosilase, TIL1e MSH2, o perfil de expressão por qRT-PCR foi similar ao identificado por RNA-Seq. Comparados com 520-S3, os níveis de transcritos desses genes foram inferiores em 520-S1 e superiores em 1404-R1. Também consistentes com os resultados do RNA-Seq foram os resultados de expressão dos genes envolvidos em resposta a estímulos abióticos, Galactose oxidase e RPT2, cujos níveis de transcritos, mensurados por qRT-PCR, foram superiores em relação a S3. Exceto por PGL1, cujo nível de transcrito por qRT-PCR foi inferior em 520-S1 quando comparado com 520-S3, os demais genes selecionados, BAR1, ADH, RSW10, GAPDH, UBQ, acumularam um

nível de transcrito determinado por qRT-PCR, em concordância com os resultados do RNA-Seq. Coletivamente, estes resultados validaram os ensaios de RNA-seq.

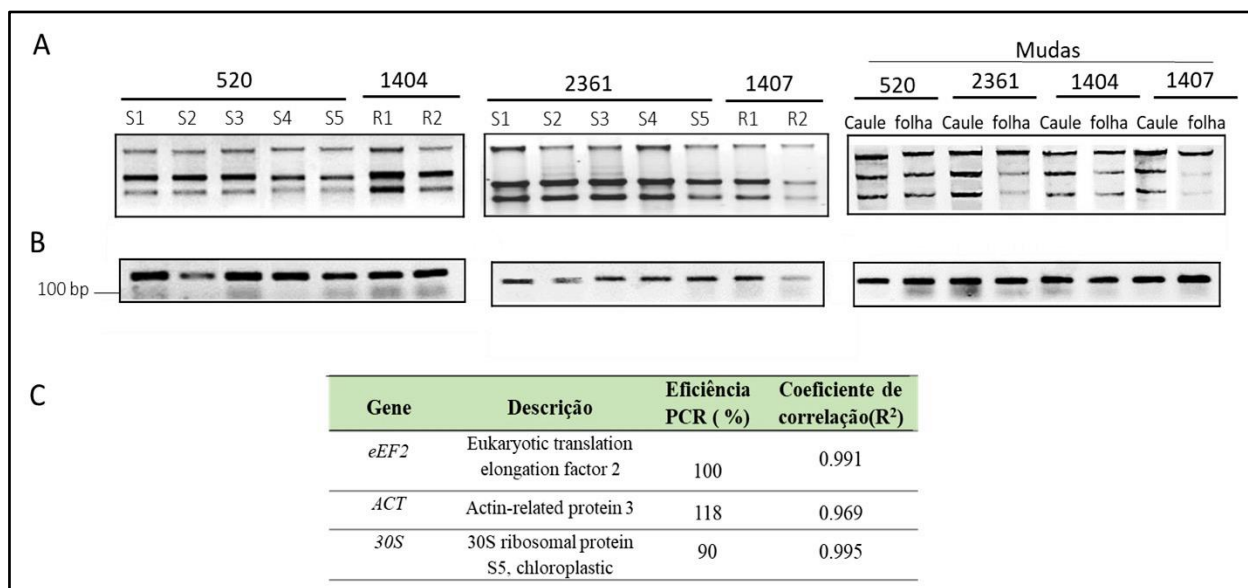


Figura 18: A) Extração de RNA total dos clones 520, 1404, 2361, 1407, e das mudas correspondentes A. Separação eletroforética de RNA extraído de caules e folhas foi realizada em gel de agarose 1% e visualizado por coloração em Brometo de Etídeo. B) Análise de RT-PCR com base no gene endógeno *eEF2*. C) Eficiência dos oligonucleotídeos utilizados como controle endógeno para reação de qRT-PCR.

Tabela 9: Genes diferencialmente expressos avaliados por qRT-PCR

Eucalipto	Gene	Descrição	Log₂ FoldChange	p-valor	Contraste	Processo biológico
Eucgr.E00458	KAS_III	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III	0.78 -0.61	0.014302789 0.037650496	R1-S1 S1-S3	Fatty acid biosynthesis
Eucgr.F00484	ACC1	acetyl-CoA carboxylase 1, synthesis of malonyl-CoA	-1.25	0.01830475	S1-S3	Fatty acid biosynthesis process
Eucgr.H00232	KCS1	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C Term	4.23 -3.68	1.52E-05 3.80E-05	R1-S1 S1-S3	Fatty acid biosynthesis elongation
Eucgr.K00927	ADH	GroES-like zinc-binding dehydrogenase family protein	1.88 -1.48	0.000750402 0.003124045	R1-S1 S1-S3	Fatty acid degradation, Glycolysis / Gluconeogenesis
Eucgr.K00013	PGL1	NagB/RpiA/CoA transferase-like superfamily protein	2.12 1.89	6.95E-05 0.000688734	R1-S3 R1-S1	Pentose phosphate pathway
Eucgr.D01028	RSW10	NagB/RpiA/CoA transferase-like superfamily protein	-1.1 0.98	0.006591108 0.006342078	R1-S1 S1-S3	
Eucgr.B03112	Glycosylase	DNA glycosylase superfamily protein	7.56 -6.07	1.56E-09 1.87E-07	R1-S1 S1-S3	Base excision repair
Eucgr.B03882	TIL1	POL2A,TIL1; DNA polymerase epsilon catalytic subunit	3.42 -2.9 -1.6	2.25E-13 5.27E-12 0.000655981	R1-S1 S1-S3 S2-S3	DNA replication, Nucleotide excision repair, Base excision repair
Eucgr.B00700	MSH2	MSH2; DNA mismatch repair protein MSH2	0.69 2.32 -1.63 -1.28	1.25E-18 1.62E-13 6.38E-12 9.06E-07	R1-S3 R1-S1 S1-S3 S2-S3	Mismatch repair
Eucgr.D00242	BARD1	ATBRCA1,BRCA1; breast cancer susceptibility1	3.56 -3.63 -1.64	3.29E-07 4.45E-09 0.022923577	R1-S1 S1-S3 S2-S3	Homologous recombination
Eucgr.B03483	Galactose oxidase	Galactose oxidase/kelch repeat superfamily protein	-5.61 5.98 3.46	0.036537475 6.3766993e-09 0.006836687	R1-S1 S1-S3 S2-S3	
Eucgr.B03325	RPT2	Root phototropism protein 2	-10.91 9.57 4.98	1.634768e-05 2.836317e-45 4.374303e-11	R1-S1 S1-S3 S2-S3	Tropismo metabolic process, Response to abiotic stimulus
Eucgr.J02288	GAPDH	glyceraldehyde 3-phosphate dehydrogenase	2.63 -2.260	0.000123073 0.00027011	R1-S1 S1-S3	Carbon fixation in photosynthetic organisms
Eucgr.K00756	Ubiquitin	small ubiquitin-like modifier 2	-1.00 -0.93	1.368476e-07 2.114482e-06	R1-S1 R1-S3	Ubiquitin mediated proteolysis

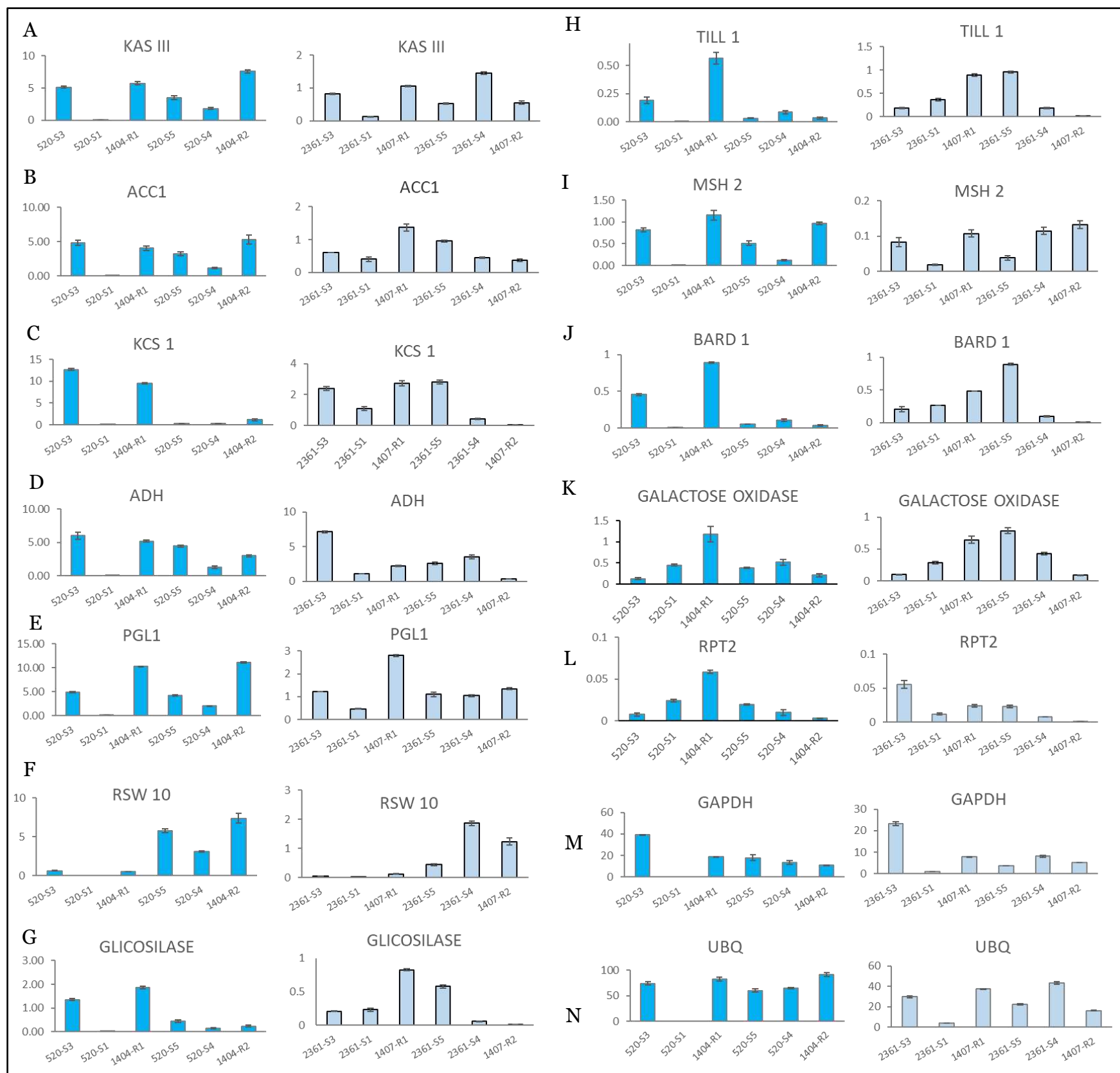


Figura 19: Acúmulo de transcritos de genes marcadores em caules sintomáticos do clone 0520 e em caules resistentes do clone 1404. O RNA total foi isolado de amostras S1, amostras S2, amostras S3 (controle) e amostras R1; a expressão dos genes indicados na figura foi determinada por qRT-PCR. São evidenciados quatorze genes marcadores representados pelas letras de A à N. Os Gráficos em azul escuro indicam análises de expressão a partir dos clones 0520 e 1404, ao passo que gráficos em azul claro indicam análises com os clones 2361 e 1407. A expressão gênica foi calculada pelo método do $2^{-\Delta CT}$; *eEF2* foi utilizado como controle endógeno. As barras indicam o desvio padrão da média, baseado em repetições técnicas e replicatas biológicas (n=3).

Com o objetivo de se avaliar se os genes diferencialmente expressos no caule também teriam o mesmo padrão de expressão nas folhas assintomáticas e sintomáticas do clone 520 e 1404, foram feitas análises de expressão por qRT-PCR dos genes selecionados nas amostras 520-S5, 520-S4 e 1404-R2. Os resultados de folhas foram consistentes com os resultados do caule, para os genes da síntese de ácidos graxos, KASIII e ACC1, para os genes de reparo do DNA, DNA RSW 10 e Glycosilase, os genes ADH, PGL1, MSH2. Os demais genes mostraram padrões de expressão nas folhas diferenciados do caule, indicando um nível de tecido-especificidade de expressão que sobrepõem à resposta típica ao distúrbio fisiológico (Figura 19).

Foram também analisados o nível de expressão dos genes selecionados em amostras sintomáticas e assintomáticas de caules e folhas do clone suscetível 2361 e do clone resistente 1407 (Figura 19). A expressão dos genes envolvidos na síntese de ácido graxo nas amostras de caules dos clones 2361 e 1407 foram similares às amostras correspondentes dos clones 520 e 1404, confirmando repressão significativa da síntese de ácidos graxos em amostras de clones suscetíveis acometidas pelo distúrbio fisiológico. Estes resultados não se confirmaram para as amostras de folhas dos clones 2361 e 1407. Também não houve perfeita concordância entre os clones suscetíveis e resistentes com os genes envolvidos no reparo do DNA, DNA Glycosilase, MSH2 e TILL1. Uma melhor correlação no nível de expressão dos genes dos diferentes clones foi observada nas amostras de caule, como ADH, RSW10, Galactose oxidase, GAPDH, UBQ. O gene PGL1 mostrou similar perfil de expressão para as amostras de folhas e caules dos clones suscetíveis e resistentes (Figura 19).

Estudos posteriores serão necessários para identificação de possíveis marcadores moleculares para resistência e suscetibilidade ao “distúrbio fisiológico” a partir da análise de expressão de genes marcadores do distúrbio em estágio de desenvolvimento de mudas.

4. DISCUSSÃO

Diagnóstico molecular do estado de “distúrbio fisiológico”

Nos últimos anos, tem havido grande expansão das plantações de eucalipto para suprir a crescente demanda mundial de matéria-prima para as indústrias de base florestal, a preferência pela cultura, se explica, devido as características que apresentam. Paralelamente à expansão das áreas plantadas, a ocorrência de um fenômeno diagnosticado como “distúrbio fisiológico” no eucalipto, tem sido relativamente frequente nas regiões sul e leste da Bahia. Este evento está diretamente relacionado com a perda de dominância apical levando à necrose do tronco principal, que se torna quebradiço, comprometendo, deste modo, a produção da cultura e principalmente a qualidade da madeira. Paralelamente, ocorre emergência de um tronco secundário e os sintomas são também associados com perda de turgor, encarquilhamento e amarelecimento das folhas. Mecanismos epigenéticos têm implicado na regulação da expressão de genes relacionados ao estresse (Chinnusamy e Zhu 2009), com o surgimento de um caule emergente logo após necrose do caule principal, porém, os mecanismos moleculares envolvidos no desenvolvimento dos sintomas e na resposta de defesa da planta ao distúrbio são totalmente desconhecidos.

A fim de entender as bases moleculares que fundamentam o estado de distúrbio fisiológico, foi conduzido um estudo de variação global de expressão gênica com amostras do clone suscetível 520, correspondentes a caules sintomáticos (S1), caules emergentes (S2) e caules assintomáticos (S3), obtidos de plantas crescidas no mesmo talhão, sob as mesmas condições ambientais. Nestas condições, uma comparação direta entre o transcriptoma de caules sintomáticos (amostra 520-S1) e do caule suscetível assintomático (amostras 520-S3) resultou na identificação de uma intensiva reprogramação gênica induzida pelo estado de “distúrbio fisiológico”, enquanto que nos caules emergentes exibiram um número muito inferior de genes diferencialmente expressos. Estes resultados indicam que as amostras S2 se encontram em recuperação apresentando um transcriptoma mais próximo das amostras assintomáticas S3.

O transcriptoma induzido pelo “distúrbio fisiológico” divergiu consideravelmente das amostras suscetíveis assintomáticas do clone 520. Nas modificações positivas de expressão

gênica induzidas pelo distúrbio fisiológico, observou-se um enriquecimento de genes DE nas categorias biológicas do sistema antioxidativo (catálase, peroxidase, desidrogenases, superóxido dismutase), morte celular (caspases, VPE, senescence-associated genes) e síntese de flavonoide. Sistemas de desintoxicação foram desenvolvidos pelas plantas como mecanismo de defesa antioxidante para combater os efeitos deletérios de ROS. Plantas com níveis mais altos de antioxidantes, apresentam maior resistência a diferentes tipos de estresse ambiental (Hasanuzzaman et al., 2012). Também foram enriquecidos nas modificações positivas de expressão gênica, genes regulatórios, normalmente envolvidos em respostas típicas a estresses abióticos, como genes da família NAC, APE, WRKY de transfatores transcricionais (Guillaumie et al., 2010). Este conjunto de genes *up* regulados categorizam as respostas majoritárias de plantas a estresses abióticos. De fato, estresses abióticos primários, resultante de mudança de temperatura, salinidade, seca e alagamento, levam a estresse osmótico secundário como efeito comum, que aciona o sistema oxidativo de plantas pelo acúmulo e produção de espécies reativas de oxigênio (ROS) (Gill & Tuteja 2010). Condições severas de estresse abióticos rapidamente induzem um conjunto de genes associados com morte celular programada em plantas, como uma última tentativa de sobrevivência das células (Costa et al., 2008; Mendes et al., 2013; Carvalho et al., 2015). Em condições de estresse prolongados, o estado de PCD pode evoluir para necrose, caracterizado pelos sintomas observados em caules acometidos pelo “distúrbio fisiológico”.

Consistente com estas observações, genes envolvidos na via de sinalização de morte celular foram altamente induzidos nos troncos sintomáticos e necróticos (S1) em relação ao controle (S3). As principais proteínas expressas em S1, envolvidas com a morte da célula incluem *N-rich proteins* (proteínas NRP), responsável por uma regulação positiva de NAC81, um efetor da resposta de morte celular, na cultura da soja (Reis & Fontes, 2012), genes NAC, relacionados a processos de desenvolvimento, sinalização auxina, defesa, respostas ao estresse abiótico e senescência foliar (Olsen, 2005; Guo, 2006), proteínas peptidase (domínio caspase) (Hatsugai et al., 2009; Lam E, 2000; Rojo E, 2004), que são consideradas efetores centrais de apoptose (Zhaoyu Jin & Wafik & El-Deiry., 2005). Além desses, também se destaca em S1 a enzima de processamento vacuolar VPE que exhibe atividade de caspase-1-like (Hatsugai et al., 2009; Rojo E, 2004). Deste modo, nossos resultados indicam que, o denominador comum de todos os sintomas do distúrbio, induzido por elementos climáticos,

parece ser a ativação descontrolada de vias de sinalização de morte celular, levando à necrose ou senescência.

No perfil do transcriptoma reprimido pelo “distúrbio fisiológico”, genes envolvidos no sistema de reparo de DNA (DNA glycosylase, TILL1, MSH2) foram extensivamente reprimidos. As plantas são particularmente vulneráveis a fatores que danificam DNA. Como defesa, desenvolveram mecanismos sofisticados para detecção de danos no DNA e reparo, dedicados a assegurar a estabilidade genômica por meio de remoção das lesões no DNA e reconstituição da informação genética original (Kireva et al., 2013). Genes da família DNA glycosylase, permitem às plantas o desenvolvimento de um mecanismo único de desmetilação ativa do DNA, pela reversão da metilação do DNA no carbono 5 da citosina (5-meC), através de uma via de reparo de excisão de base (BER) (Parrilla-doblas et al., 2019). A desmetilação ativa do DNA neutraliza a metilação excessiva, e uma de suas principais funções é evitar que a metilação se espalhe para genes próximos (Parrilla-doblas et al., 2019). As proteínas DNA-glycosylases, também estão envolvidas em respostas das plantas a uma variedade de estímulos ambientais bióticos e abióticos. Logo, se o sistema de reparo de DNA é comprometido, mecanismo de desmetilação se inviabiliza, e instabilidade genômica gera um sinal para ativação de vias de morte celular. Conseqüentemente, a instabilidade do genoma causada por um sistema ineficiente de reparo do DNA, em condições de estresses abióticos, pode ser o sinal de ativação das vias de morte celular induzida pelo distúrbio fisiológico.

Também relevante foi a repressão expressiva dos genes envolvidos na síntese de ácidos graxos, incluindo ácidos graxos de cadeia longa e ceras, nas amostras 520-S1. Estes genes, estão envolvidos em diversas etapas de desenvolvimento e crescimento de plantas, como síntese de ceras e cutina nos caules e folhas, formação da raiz lateral, controle de divisão celular na região de meristemas (Bonaventure et al., 2003). Yaeno e colaboradores em 2004, mostraram que os ácidos graxos trienólicos (ATs) nas membranas dos cloroplastos, estão envolvidos nas respostas de defesa contra patógenos bacterianos avirulentos. A invasão de patógenos em plantas induz uma produção de ROS, e subsequente morte celular programada. Desde modo, reduções ou bloqueios na síntese de ácidos graxos saturados têm sido associados a declínios na taxa de crescimento de plantas, na síntese de ceras e cutina, um dos principais componentes da cutícula da planta e indução de morte celular (Bonaventure et

al.,2003; Baud et al., 2003; Schnurr & Shockey, 2004). Embora todas as mutações nos genes da biossíntese de ácidos graxos tenham sido geradas e analisadas em *Arabidopsis*, o fenótipo de plantas de eucalipto exibindo “distúrbio fisiológico” pode ser amplamente explicado pela redução na expressão de genes essenciais na biossíntese de ácidos graxos de cadeia longa.

Mecanismo de resistência ao distúrbio fisiológico

Para análise do possível mecanismo de resistência ao distúrbio fisiológico foi analisado os genótipos contrastantes, clone suscetível 520 e resistente 1404, crescidos sob condições climáticas idênticas e no mesmo talhão. Portanto, os transcriptomas diferenciais entre as amostras de caules assintomáticas do clone 520 e do clone 1404, refletem um somatório do efeito do genótipo e da resistência ao distúrbio fisiológico. Na tentativa de eliminar o efeito de genótipo foi analisado genes diferencialmente expressos no clone 1404 (1404-R1 vs 530-S3) em direções opostas àqueles diferencialmente expressos pelo distúrbio fisiológico (520-S1 vs 520-S3). Assim sendo, genes altamente reprimidos pelo “distúrbio fisiológico”, mas não alterados nas amostras R1 foram considerados resultantes do estado de “distúrbio fisiológico” e explicam os sintomas observados. Nesta classe, incluem os genes envolvidos na síntese de ácidos graxos. Genes induzidos nas amostras 1404-R1 mas não alterados pelo distúrbio fisiológico foram considerados resultantes de variações inerentes às diferenças de *background* genético dos genótipos contrastantes. Nesta classe, incluem genes da maquinaria de tradução das células vegetais. Por outro lado, genes *down* regulados pelo distúrbio fisiológico e *up* regulados nos clones 1404 foram considerados como participantes do mecanismo de resistência. Exemplos nesta categoria incluem os genes envolvidos no mecanismo de reparo do DNA. Para avaliar a consistência dessa interpretação, foram realizados estudos de expressão gênica por qRT-PCR no clone 2361 suscetível e 1407 resistente. Os três genes do sistema de reparo do DNA avaliados apresentaram o mesmo perfil de expressão dos clones 520 e 1404, substanciando o argumento de que a expressão desses genes poderia ser avaliada como marcador de resistência.

Na análise dos mecanismos de resistência ao distúrbio fisiológico, foi observado que genes envolvidos no sistema imune inato da planta, relacionados com PTI (*PAMP-triggered immunity*) e ETI (*effector-triggered immunity*), foram extensivamente *down* regulados nas

amostras resistentes (R1), sugerindo que imunidade inata de plantas não está positivamente associada à resistência ao “distúrbio fisiológico” nos clones resistentes (1404). Estes resultados também demonstram que o estímulo principal associado ao distúrbio fisiológico não se correlaciona a sinais bióticos. Consistente com esta interpretação, estudos de metagenômica em todas as amostras analisadas não foi capaz de delimitar predominância de patógenos nas amostras de caules.

Em resumo, os estudos de variação global de expressão gênica aqui descritos, apresentaram um cenário consistente com os fenótipos observados (sintomas) pelo estabelecimento do “distúrbio fisiológico”. Embora os resultados deste trabalho contribuam para o direcionamento das causas do distúrbio, pesquisas adicionais são necessárias para concretização efetiva dos genes que regulam as vias moleculares que eventualmente levam à ativação da morte celular durante o fenômeno do distúrbio. Assim, estudos adicionais são necessários para o desenvolvimento de marcadores moleculares como forma de detecção da predisposição genética de clones ao “distúrbio fisiológico”. A identificação do transcriptoma induzido pelo “distúrbio fisiológico” em folhas são requeridos para um diagnóstico molecular completo do distúrbio fisiológico.

5. REFERÊNCIAS

- ABRAF.** Anuário Estatístico da ABRAF de 2012 - ano base 2011. Brasília. Associação Brasileira de Produtores de Florestas Plantadas, 7, 150p., 2012.
- Alfenas AC, Zauza EAV, Mafia RG, Assis TF.** (2009) Clonagem e doenças do eucalipto. 2ª Ed. Editora UFV. Viçosa, MG. 500 p.
- Almeida, Catarina Faria de.** Study of the role of SETD2 mutations in clear cell renal cell carcinoma (ccRCC). 2013. Tese de Doutorado.
- Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D.** (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols*, 8(9), 1765.
- Anders, S., Reyes, A., & Huber, W.** (2012). Detecting differential usage of exons from RNA-seq data. *Genome research*, gr-133744.
- Associação brasileira de celulose e papel.** (2000). Distribuição geográfica da produção em toneladas. Relatório Estatístico BRACELPA. São Paulo, 2000. p.204-205.
- Auer, C. G.** (1996). A ocorrência do cancro do eucalipto no estado do Paraná e de Santa Catarina. *Boletim de Pesquisa Florestal, Colombo*, v. 32/33, p. 81-85.
- Auer, C. G.; Santos, A. F. dos; Bora, K. C.** (2010). A ferrugem do eucalipto na região Sul do Brasil. Colombo: Embrapa Florestas. 5 p. (Embrapa Florestas. Comunicado técnico, 252).
- Batoko, Henri, et al.** (2017). Understanding and exploiting autophagy signaling in plants. *Essays in biochemistry* 61.6: 675-685.
- Barba, M.; Czosnek, H.; Hadidi, A.** (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*, v. 6, n. 1, p. 106-36, Jan.
- Baud, S., Guyon, V., Kronenberger, J., Wuillème, S., Miquel, M., Caboche, M., ... & Rochat, C.** (2003). Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in Arabidopsis. *The Plant Journal*, 33(1), 75-86.
- Bioinformatics Group. 3402bioinformatics:** NEXT GENERATION SEQUENCING. Disponível em : <<http://www.3402bioinformatics.com/service/>>. Acesso em: 25 de janeiro de 2020.
- Bolger A. M., Lohse M., Usadel B.** (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30,15, 2114-2120.
- Bonaventure, G., Salas, J. J., Pollard, M. R., & Ohlrogge, J. B.** (2003). Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. *The Plant Cell*,15(4), 1020-1033.

- Bray, E.A.** (1997). Plant responses to water deficit. *Trends in Plant Science*, v.2, p.48-54.
- Brooker, M. I. H.** (2000). A new classification of genus *Eucalyptus* L'Her. (*Myrtaceae*). *Australian Systematic Botany* 13: 79-148.
- Callister, A.N.; Arndt, S.K.; Adams, M.A.** (2006). Comparison of four methods for measuring osmotic potential of tree leaves. *Physiologia Plantarum*, v.127, p.383-392.
- Carvalho, H. H., Brustolini, O. J., Pimenta, M. R., Mendes, G. C., Gouveia, B. C., Silva, P. A., & Fontes, E. P.** (2014). The molecular chaperone binding protein BiP prevents leaf dehydration-induced cellular homeostasis disruption. *PloS one*,9(1), e86661.
- Carvalho, M. C. D. C. G., & Silva, D. C. G. D.** (2010). Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. *Ciência Rural*, 40(3), 735-744.
- Chapelev, I.; WEI, G.; Tang, Q.; Zhao, K.** (2009). Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Res*, v. 37, n. 106.
- Chaves, M.M.; Maroco, J.P.; Pereira, J.S.** (2003). Understanding plant responses to drought – from genes to the whole plant. *Functional Plant Biology*, v.30, p.239-264.
- CHAVES, J.H. et al. Seleção precoce de clones de eucalipto para ambientes com disponibilidade diferenciada de água no solo: relações hídricas de plantas em tubetes. *Revista Árvore*, v.28, n.3, p.333-341, 2004.
- Chang, S., Puryear, J.; Cairney, J.** (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Report*, v. 11, p. 113-116.
- Chen, Y., Lun, A. T., & Smyth, G. K.** (2014). Differential expression analysis of complex RNA-seq experiments using edgeR. In *Statistical analysis of next generation sequencing data* (pp. 51-74). Springer, Cham.
- Chinnusamy V, Zhu JK** (2009) Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol* 12:133–139
- Costa, M. D., Reis, P. A., Valente, M. A. S., Irsigler, A. S., Carvalho, C. M., Loureiro, M. E., & Fontes, E. P.** (2008). A new branch of endoplasmic reticulum stress signaling and the osmotic signal converge on plant-specific asparagine-rich proteins to promote cell death. *Journal of Biological Chemistry*,283(29), 20209-20219.
- Dell, B.; Malajczuk, N.** (1994). Boron deficiency eucalypt in China. *Canadian Journal Forest Research*, v.24, n.12, p.2409-2416.
- Eldridge, K., Davidson, J., Harwood, C. e WYK, G.** (1994). *Eucalypt Domestication and Breeding*. USA. Oxford University Press, 1, 308p.
- Falcon S e Gentleman R** (2007). "Usando GOstats para testar listas de genes para associação de termos GO." *Bioinformatics*, 23 (2), pp. 257-8.

Ferreira, F. A. (1989) Patologia florestal: principais doenças florestais no Brasil. Viçosa, MG: SIF, 570 p.

Flexas, J.; Bota, J.; Escalona, J.M.; Sampol, B.; Medrano, H. (2002). Effects of drought on photosynthesis in grapevines under field conditions. *Functional Plant Biology*, v.29, p.461-471.

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods*, 8(6), 469.

Gill, S. S., & Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant physiology and biochemistry*, 48(12), 909-930.

Guillaumie, S., Mzid, R., Méchin, V., Léon, C., Hichri, I., Destrac-Irvine, A., ... & Lauvergeat, V. (2010). The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco. *Plant molecular biology*, 72(1-2), 215.

Guo, Y., & Gan, S. (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *The Plant Journal*, 46(4), 601-612.

Grattapaglia, D., Vaillancourt, R., Shepherd, M., Thumma, B., Foley, W., Külheim, C., Potts, B. e Myburg, A. (2012). Progress in *Myrtaceae* genetics and genomics: *Eucalyptus* as the pivotal genus. *Tree Genetics & Genomes* 8: 463-508.

Grattapaglia, D., Plomion, C., Kirst, M. e Sederoff, R. R. (2009). Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* 12: 148-156.

Hasanuzzaman, M., Hossain, M. A., da Silva, J. A. T., & Fujita, M. (2012). Plant response and tolerance to abiotic oxidative stress: antioxidant defense is a key factor. In *Crop stress and its management: Perspectives and strategies* (pp. 261-315). Springer, Dordrecht.

Hara-Nishimura, I., & Hatsugai, N. (2011). The role of vacuole in plant cell death. *Cell Death & Differentiation*, 18(8), 1298-1304.

Hausman JF, Evers D, Thiellement H, Jouve L (2000) Compared responses of poplar cuttings and in vitro raised shoots to short-term chilling treatments. *Plant Cell Reports* 19: 954-960.

Hatsugai N, Iwasaki S, Tamura K, Kondo M, Fuji K, Ogasawara K et al (2009). A novel membrane fusion-mediated plant immunity against bacterial pathogens. *Genes*; **23**: 2496–2506.

Holt, R. A.; Jones, S. J. (2008). The new paradigm of flow cell sequencing. *Genome Res.*, v. 18, p. 839–846.

ILLUMINA, TRUSEQ, R. N. A. TruSeq™ RNA and DNA Sample Preparation Kits v2. CHIP-SEQ RIBO-SEQ, p. 133.

Jin, Z., & El-Deiry, W. S. (2005). Overview of cell death signaling pathways. *Cancer biology & therapy*, 4(2), 147-171.

Jin H, Cominelli E, Bailey P, Parr A, Mehrstens F, Jones J, Tonelli C, Weisshaar B, Martin C (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *EMBO J* 19: 6150-6161.

Kawaoka A, Ebinuma H (2001) Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry* 57: 1149-1157.

Kawaoka A, Nanto K, Ishii K, Ebinuma H, Seite U (2006) Reduction of lignin content by suppression of expression of the LIM domain transcription factor in *Eucalyptus camaldulensis*. *Silvae Genetica* 55: 269-277.

Kim D, Langmead B and Salzberg SL. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*.

Kim, B. M., Choi, B. S., Lee, K. W., Ki, J. S., Kim, I. C., Choi, I. Y., ... & Lee, J. S. (2013). Expression profile analysis of antioxidative stress and developmental pathway genes in the manganese-exposed intertidal copepod *Tigriopus japonicus* with 6K oligochip. *Chemosphere*, 92(9), 1214-1223.

Kireeva, M. L., Kashlev, M., & Burton, Z. F. (2013). RNA polymerase structure, function, regulation, dynamics, fidelity, and roles in gene expression.

Krugner, T. L. (1980). Doenças do eucalipto: *Eucalyptus* spp. In: GALLI, F. (Coord.). Manual de fitopatologia. São Paulo: Agronômica Ceres. v. 2. p. 275-296.

Lam E, del Pozo O (2000). Caspase-like protease involvement in the control of plant cell death. *Plant Mol Biol*; 44: 417–428.

Larcher, W. (2000). Plantas sob estresse. In: LARCHER, W. *Ecofisiologia vegetal*. São Carlos: Rima. p.341-430.

Legay S, Lacombe E, Goicoechea M, Briere C, Seguin A, Mackay J, Grima-Pettenati J (2007) Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science* 173: 542-549

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*, 41(10), e108-e108.

Love, MI, W. Huber e S. Anders. (2014). “Estimativa moderada de alteração e dispersão de dobras para dados de RNA-seq com DESeq2.” *Genome Biology* 15 (12): 550

McCarthy RL, Zhong R, Ye Z-H (2009) MYB83 is a direct target of *snd1* and acts redundantly with *myb46* in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant and Cell Physiology* 50: 1950-1964.

Mele G, Ori N, Sato Y, Hake S (2003) The knotted1-like homeobox gene BREVIPEDICELLUS regulates cell differentiation by modulating metabolic pathways. *Genes Dev* 17: 2088-2093.

Mendes, G. C., Reis, P. A., Calil, I. P., Carvalho, H. H., Aragão, F. J., & Fontes, E. P. (2013). GmNAC30 and GmNAC81 integrate the endoplasmic reticulum stress-and osmotic stress-induced cell death responses through a vacuolar processing enzyme. *Proceedings of the National Academy of Sciences*, 201311729.

Exportação de produtos florestais cresce e segmento já é o segundo da pauta do agro. **Ministério da Agricultura**, 2018. Disponível em: <<http://www.agricultura.gov.br/noticias/exportacao-de-produtos-florestais-cresce-e-segmento-ja-e-o-segundo-da-pauta-do-agro>>. Acesso em: 30 de janeiro de 2020.

Moura JC, Bonine CA, Viana JO, Dornelas MC, Mazzafera P (2010) Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J Integr Plant Biol* 52: 360-376

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., & Goodstein, D. M. (2014). The genome of *Eucalyptus grandis*. *Nature*, 510(7505), 356-362.

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., & Goodstein, D. M. (2014). The genome of *Eucalyptus grandis*. *Nature*, 510(7505), 356-362. 1993.

Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends in plant science*, 10(2), 79-87.

Osório, J.; Osório, M.L.; Chaves, M.M.; Pereira, J.S. (1998). Water deficits are more important in delaying growth than in changing patterns of carbon allocation in *Eucalyptus globulus*. *Tree Physiology*, v.18, p.363- 373.

Paiva JA, Prat E, Vautrin S, Santos MD, San-Clemente H, Brommonschenkel S, Fonseca PG, Grattapaglia D, Song X, Ammiraju JS, Kudrna D, Wing RA, Freitas AT, Bergès H, Grima-Pettenati J (2011) Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. *BMC Genomics* 12: 137.

Parrilla-Doblas, J. T., Roldán-Arjona, T., Ariza, R. R., & Córdoba-Cañero, D. (2019). Active DNA demethylation in plants. *International journal of molecular sciences*, 4683.

Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, Campbell MM (2003) Characterisation of a pine MYB that regulates lignification. *Plant Journal* 36: 743-754.

Pimenta, M. R., Silva, P. A., Mendes, G. C., Alves, J. R., Caetano, H. D. N., Machado, J. P. B., ... & Rosado, G. L. (2016). The stress-induced soybean NAC transcription factor GmNAC81 plays a positive role in developmentally programmed leaf senescence. *Plant and Cell physiology*, 57(5), 1098-1114.

Pimentel, C. (2005). Respostas fisiológicas à falta d'água: limitação difusiva ou metabólica? In: NOGUEIRA, R.J.M.C.; ARAÚJO, E. DE L.; WILLADINO, L.G.; CAVALCANTE, U.M.T.; (Ed.). *Estresses ambientais: danos e benefícios em plantas*. Recife: UFRPE, Imprensa Universitária, p.13-21.

Poke, F. S., Vaillancourt, R. E., Potts, B. M. e Reid, J. B. (2005). Genomic Research in *Eucalyptus*. *Genetica* 125: 79-101.

Reis, P. A., Rosado, G. L., Silva, L. A., Oliveira, L. C., Oliveira, L. B., Costa, M. D, & Fontes, E. P. (2011). The binding protein BiP attenuates stress-induced cell death in soybean via modulation of the N-rich protein-mediated signaling pathway. *Plant physiology*, pp-111.

Rezende GDSP, Resende MDV, Assis TF 2014. Eucalyptus breeding for clonal forestry. In: Fenning T, ed. Challenges and opportunities for the World's forests in the 21st century. Dordrecht, the Netherlands: Springer, 393–424.

Reis, P. A., & Fontes, E. P. (2012). N-rich protein (NRP)-mediated cell death signaling: a new branch of the ER stress response with implications for plant biotechnology. *Plant signaling & behavior*, 7(6), 628-632.

Rocha Filho, J.V.C.; Haag, H.P.; Oliveira, G.D. (1978). Deficiência de macronutrientes, boro e ferro em *Eucalyptus urophylla*. Anais da Escola Superior de Agricultura “Luiz de Queiroz”, v.35, p.19-34.

Rogers, H. J. (2015). Senescence-associated programmed cell death. In *Plant Programmed Cell Death* (pp. 203-233). Springer, Cham.

Rojo E, Martin R, Carter C, Zouhar J, Pan S, Plotnikova J et al (2004). VPEgamma exhibits a caspase-like activity that contributes to defense against pathogens. *Curr Biol*; **14**: 1897–1906.

Savory, B.M. (1962). Boron deficiency in *Eucalyptus* in Northern Rhodesia. *Empire Forest Review*, v.41, n.2, p.118-125.

Schnurr, J., & Shockey, J. (2004). The acyl-CoA synthetase encoded by LACS2 is essential for normal cuticle development in *Arabidopsis*. *The Plant Cell*, 16(3), 629-642.

Seki, M.; Umezawa, T.; Urano, K.; Shinozaki, K. (2007). Regulatory metabolic networks in drought stress responses. *Current Opinion in Plant Biology*, v.10, p.296:302.

Sharma, P., Jha, A. B., Dubey, R. S., & Pessarakli, M. (2012). Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. *Journal of botany*, 2012.

Sheth, Bhavisha P.; Thaker, Vrinda S (2014). Plant systems biology: insights, advances and challenges. *Planta*, v. 240, n. 1, p. 33-54.

Silveira, R.L.V.A.; Takahashi, E.N.; Sgarbi, F.; Branco, E.F. (1996). Sintomas de deficiência de macronutrientes e boro em híbrido de *Eucalyptus grandis* x *Eucalyptus urophylla* em solução nutritiva. In: CONGRESSO LATINO AMERICANO DE CIÊNCIA DO SOLO, 13., Anais (compact disc). Águas de Lindóia: Sociedade Americana de Ciência do Solo.

Silveira, R. L. V. D. A., Casarin, V., Paula, T. D. A., & Silveira, R. I. (2007). Nutrição e adubação com Boro em *Eucalyptus*. *Boletim Técnico*. Piracicaba: Agroflorestal.

Silva, M.R. (1998). Caracterização morfológica, fisiológica e nutricional de mudas de *Eucalyptus grandis* Hill ex Maiden submetidas a diferentes níveis de estresse hídrico durante a fase de rustificação. 105f. Dissertação (Mestrado e Silvicultura) - Universidade do Federal do Paraná, Curitiba, 1998.

Soewarto, J., Hamelin, C., Bocs, S., Mournet, P., Vignes, H., Berger, A., & Carriconde, F. (2019). Transcriptome data from three endemic Myrtaceae species from New Caledonia displaying contrasting responses to myrtle rust (*Austropuccinia psidii*). *Data in brief*, 22, 794-811.

Souza, J.R., & Ribeiro, B.N. Estresses Abióticos! Como Conviver com Eles?. **Agrolink**, 2015. Disponível em: <https://www.agrolink.com.br/colunistas/estresses-abioticos--como-conviver-com-eles--_387950.html>. Acesso em 31 de janeiro de 2020.

Souza, M.J.H.; Ribeiro, A.; Leite, H.G.; Leite, F.P.; Minuzzi, R.B. (2006). Disponibilidade hídrica do solo e produtividade do eucalipto em três regiões da Bacia do Rio Doce. *Revista Árvore*, v.30, p.399-410.

Souza, U. R. (2013). Fatores de transcrição relacionados à lignina e resposta ao estresse hídrico e à baixa temperatura em eucalipto.

Stone, E.L. (1990). Boron deficiency and excess in forest trees: a review. *Forest Ecology and Management*, v.37, n.1/3, p.49-75.

Taiz, L.; Zeiger, E. (2004). Fisiologia do estresse. In: SANTAREM, E. R. et al. *Fisiologia vegetal*. 3.ed. Porto Alegre: Artmed. p.613-641.

Taylor, I.B.; Burbidge, A.; Thompson, A.J. (2000). Control of abscisic acid synthesis. *Journal of Experimental Botany*, v.51, p.1563-1575.

Tokeshi, F.; Guimarães, R.F.; Tomazello Filho, M. (1976). Deficiência de boro em *Eucalyptus* em São Paulo. *Summa Phytopathologica*, v.2, n.2, p.122-126.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), 562-578., v. 7, n. 3, p. 562-578, 2012.

Vail, J.W.; Calton, W.E.; Strang, R.M. (1957). Dieback of Wattle: boron deficiency. *Empire African Agriculture Journal*, v.23, n.2, p.100-103.

Varuzza, L. (2013). Introdução à análise de dados de sequenciadores de nova geração. **Apostila de bioinformática, 76p.**

Wang H, Zhao Q, Chen F, Wang M, Dixon RA (2011) NAC domain function and transcriptional control of a secondary cell wall master switch. *Plant J* 68: 1104-1114

Wang HZ, Dixon RA (2012) On-off switches for secondary cell wall biosynthesis. *Mol Plant* 5: 297-303

Wang, Z.; Gerstein, M.; Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, v. 10, p. 57-63.

Williams, J. E. e Woinarski, J. (1997). *Eucalypt ecology: Individuals to ecosystems*. Cambridge University Press 430p.

Wright, I.J.; Groom, P.K.; Lamont, B.B.; Poot, P.; Prior, L.A.; Reich, P.B.; Schulze, ED.; Veneklaas, E.J.; Westoby, M. (2004). Leaf trait relationships in Australian plant species. *Functional Plant Biology*, v.31, p.551-558.

Yaeno T, Matsuda O, Iba K. (2004). Role of chloroplast trienoic fatty acids in plant disease defense responses. *Plant J.* 40:931–41

Yamaguchi M, Mitsuda N, Ohtani M, Ohme-Takagi M, Kato K, Demura T (2011) VASCULAR-RELATED NAC-DOMAIN7 directly regulates the expression of a broad range of genes for xylem vessel formation. *Plant J* 66: 579-590.

Zhang DW, Shao J, Lin J, Zhang N, Lu BJ, Lin SC et al (2009). RIP3, an energy metabolism regulator that switches TNF-induced cell death from apoptosis to necrosis. *Science*; **325**: 332–336.

Zhang, K., & Gan, S. S. (2012). An abscisic acid-AtNAP transcription factor-SAG113 protein phosphatase 2C regulatory chain for controlling dehydration in senescing Arabidopsis leaves. *Plant Physiology*, 158(2), 961-969.

Zhong R, Ye Z-H (2007) Regulation of cell wall biosynthesis. *Current Opinion in Plant Biology* 10: 564-572.

Zhong R, Lee C, Zhou J, McCarthy RL, Ye Z-H (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell* 20: 2763-2782.

Zhong R, Ye ZH (2009) Transcriptional regulation of lignin biosynthesis. *Plant Signal Behav* 4: 1028-1034.