

PROSPECÇÃO E ANOTAÇÃO DE GENES RESPONSIVOS A ESTRESSE DE SALINIDADE EM DENDÊ (*Elaeis guineensis* Jacq.)

Thalita Massaro Malheiros Ferreira¹, André Pereira Leão², Carlos Antonio Ferreira de Sousa³,
Manoel Teixeira Souza Júnior⁴

RESUMO: O RNA-seq é técnica mais utilizada atualmente para caracterização do transcriptoma de um organismo, e se baseia no sequenciamento de cDNAs derivados de transcritos, em larga escala, utilizando plataformas de NGS. A análise de dados de RNA-seq permite identificar genes diferencialmente expressos em uma determinada condição, como o estresse salino. Este trabalho visou à prospecção de genes de *E. guineensis* responsivos a estresse de salinidade, e sua caracterização funcional e estrutural. Folhas de plantas de dendê submetidas a duas doses distintas de NaCl (0.0 e 2.0g), coletadas aos cinco e 12 dias após início do estresse, foram submetidas a extração de RNA total (três plantas por tratamento). A plataforma de bioinformática OmicsBox, na sua versão 1.2.4., foi empregada na seleção e anotação dos genes responsivos ao estresse salino, e para gerar o perfil de expressão *in silico* dos genes selecionados. O genoma completo de *E. guineensis* disponível no NCBI foi utilizado como referência - BioProject PRJNA192219. Foram identificados 33 genes diferencialmente expressos quando comparando cinco dias de estresse e o controle, e dez genes quando comparando 12 dias de estresse e o controle; utilizando os seguintes critérios de seleção: $FDR \leq 0.01$ e $\log FC \geq 5$. Seis genes eram comuns aos dois grupos e foram escolhidos para a anotação funcional e estrutural.

PALAVRAS-CHAVE: estresse abiótico, palma de óleo, transcritômica

PROSPECTION AND ANNOTATION OF GENES RESPONSIVE TO OIL PALM SALINITY STRESS (*Elaeis guineensis* Jacq.)

ABSTRACT: RNA-seq is the technique most used today to characterize an organism's transcriptome and is based on the large-scale sequencing of transcript-derived cDNAs using

¹ Doutoranda, PGBV, Univ. Federal de Lavras (UFLA), CEP 37200-000, Lavras, MG, Fone (61).99259.1869, thatamassaro95@gmail.com

² M.Sc., Embrapa Agroenergia, CEP 70770-901, Brasília, DF

³ D.Sc., Embrapa Meio Norte, Caixa Postal 001, CEP 64008-780 Teresina, PI

⁴ Ph.D., Embrapa Agroenergia, CEP 70770-901, Brasília, DF

NGS platforms. The analysis of RNA-seq data allows us to identify genes differentially expressed in a given condition, such as saline stress. This work aimed to search for *E. guineensis* genes responsive to salinity stress and their functional and structural characterization. Leaves of oil palm plants submitted to two different doses of NaCl (0.0 and 2.0g per 100g of the substrate), collected at five and 12 days after the beginning of the stress, were subjected to extraction of total RNA (three plants per treatment). The OmicsBox bioinformatics platform, version 1.2.4., was applied in the selection and annotation of genes responsive to salt stress and to generate the *in silico* expression profiles. The complete genome of *E. guineensis*, available at NCBI, was used as a reference - BioProject PRJNA192219. A total of 33 differentially expressed genes were found when comparing five days of stress and control and ten genes when comparing 12 days and control; using the following selection criteria: $FDR \leq 0.01$ and $\log FC \geq 5$. Six genes were common to both groups and chosen for functional and structural annotation.

KEYWORDS: abiotic stress, oil palm, transcriptomics

INTRODUÇÃO

Transcritômica é o estudo do transcrito - o conjunto completo de RNA, também conhecido como perfil de expressão, que são produzidos pela transcrição do genoma, sob circunstâncias específicas ou em uma célula, um tecido, um órgão, ou do organismo inteiro, em um determinado momento do desenvolvimento deste. Diversos tipos de RNA são encontrados no transcrito; tais como: mRNA, tRNA, rRNA, snRNA, snoRNA, miRNA, lncRNA e pseudogenes (WANG et al., 2017). Nos últimos anos, em decorrência dos avanços tecnológicos alcançados com a técnica de RNA-seq, e da significativa redução dos custos para aplicação da mesma, testemunhamos uma explosão na quantidade de dados de transcrito gerados e tornados públicos (LOWE et al., 2017).

RNA-seq se refere ao sequenciamento em larga escala de cDNAs derivados de transcritos, mediante emprego de tecnologias de sequenciamento de nova geração, sendo que a abundância desses é derivada do número de “reads” para cada transcrito (LOWE et al., 2017). A análise dos dados de RNA-seq se inicia com uma análise da qualidade das sequências geradas, seguida por uma etapa de pré-processamento para melhorar a qualidade destas, e o mapeamento das sequências de alta qualidade em um genoma ou transcrito de referência. Por fim, os dados mapeados são normalizados e utilizados para análise de expressão diferencial (ED) de genes mediante emprego de métodos estatísticos e de

aprendizado de máquina (“*machine learning*”) Costa-Silva et al. (2017) citado por Stark et al. (2019).

A análise de ED é provavelmente a mais importante etapa que permite ao usuário interpretar as informações biológicas sobre os perfis de expressão de uma célula, um tecido, um órgão, ou do organismo inteiro, em um determinado momento do desenvolvimento deste, em determinadas condições ambientais distintas (JIMÉNEZ-JACINTO et al., 2019). O perfil de expressão gênica *in silico* obtido com a análise de ED contém as informações sobre genes relacionados à resposta do organismo a uma determinada condição como, por exemplo, estresse salino.

A salinidade do solo é um problema presente em mais de 100 países; sendo que aproximadamente 20% das terras agriculturáveis no mundo apresentam solos salinos e/ou sódicos, e entre 25% e 30% das terras irrigadas são afetadas pelo sal e essencialmente improdutivas comercialmente (SHAHID et al., 2018). A agricultura bioassalina é uma forma relativamente nova de lidar com a salinidade na agricultura; e, neste contexto, são desenvolvidos sistemas de cultivo para ambientes salinos, utilizando a capacidade de certas plantas para crescer sob condições salinas em combinação com o uso de solos e recursos hídricos salinos, e um melhor manejo do solo e da água. A agricultura bioassalina pode ser implementada com espécies glicófitas tolerantes à salinidade ou com halófitas; e essas duas possibilidades não são excludentes (BORSALINI et al., 2018). O uso de glicófitas pode se materializar a partir de duas estratégias distintas, e também não excludentes; são elas: a) coleta, caracterização, e domesticação de espécies selvagens de glicófitas tolerantes à salinidade; e b) transferência vertical e/ou horizontal da tolerância à salinidade - encontrada em halófitas e em glicófitas - para as principais culturas agrícolas atualmente utilizadas (e que são na sua maioria glicófitas não tolerantes à salinidade) (BRAGA et al., In press).

Vieira et al. (2020) avaliou as respostas morfofisiológicas e o desequilíbrio iônico no substrato, raízes e folhas de plantas jovens dendê (*Elaeis guineensis* Jacq.) sob diferentes níveis de estresse salino, seguindo um protocolo de salinização de substrato no qual o nível de sal era conhecido. Este presente estudo utilizou amostras coletadas por Vieira et al. (2020) e submetidas a RNA-seq, e objetivou prospectar genes de *E. guineensis* responsivos a estresse de salinidade e realizar a sua anotação estrutural e funcional.

MATERIAL E MÉTODOS

Plantas jovens de dendê foram submetidas a duas doses distintas de NaCl (0.0 e 2.0 g de NaCl por 100 g de substrato), e mantidas nestas condições por 12 dias. Folhas de três plantas estressadas foram coletadas aos 5 e 12 dias após o início do estresse, juntamente com folhas de 3 plantas controle (aos 12 dias), imediatamente congeladas em nitrogênio líquido, e estocadas a -80°C até a extração de RNA total. O Qiagen RNeasy® Plant Mini kit (QIAGEN, CA, USA) foi utilizado para a extração do RNA total, seguindo orientação do fabricante, e enviadas para a empresa GenOne (Rio de Janeiro, Brasil), para produção das bibliotecas e sequenciamento, pair-end, 150 nucleotídeos por read. Os arquivos fastq com as sequências foram depositadas no “Sequence Read Archive” (SRA) database do “National Center for Biotechnology Information” - BioProject PRJNA573093, BioSample SAMN12799239.

Para a análise da expressão diferencial visando à prospecção de genes responsivos à salinidade em dendê, foi utilizada a plataforma de bioinformática OmicsBox (OmicsBox – Bioinformatics Made Easy, BioBam Bioinformatics, March 3, 2019, <https://www.biobam.com/omicsbox>), na sua versão 1.2.4., módulo “transcritomics”, utilizando o “Reference Based Transcriptomics Workflow” ajustado.

Os 18 arquivos fastq gerados a partir 9 amostras sequenciadas foram submetidos ao pré-processamento, e posterior análise de qualidade. Por fim, os arquivos com as sequências de alta qualidade foram submetidos ao mapeamento contra o genoma completo de *Elaeis guineensis* disponível no NCBI foi utilizado como referência - BioProject PRJNA192219. Os 9 arquivos BAM gerados durante a etapa de mapeamento foram submetidos à geração da Count Table (BAM+GFF) utilizando o “software HTSeq”, e análise de expressão diferencial “pairwise” utilizando o “software package edgeR”.

Para selecionar os genes responsivos à salinidade, foram utilizados os seguintes critérios: $FDR \leq 0.01$ e $\log_{2}FC \geq 5$. Estes critérios foram utilizados para selecionar genes diferencialmente expressos entre as plantas controle e as estressadas, tanto aos 5 quanto aos 12 dias de estresse. Os genes selecionados como *up regulated* foram submetidos à anotação estrutural e funcional, tendo como base o genoma de dendê disponível no NCBI (SINGH et al., 2013).

RESULTADOS E DISCUSSÃO

Após o pré-processamento e análise de qualidade foi feito o mapeamento, que remete ao processo de alinhamento das sequências das amostras contra o genoma de referência de dendê (Tabela 1). Após o mapeamento, foi gerada uma tabela de contagem que estima a expressão gênica de experimentos de sequenciamento de RNA. Essa ferramenta utiliza arquivos com leituras de sequenciamento alinhadas no formato SAM/BAM e um arquivo GTF/GFF com coordenadas de características genômicas. Ele faz a contagem de quantas leituras são mapeadas para cada recurso de interesse (genes, exon). A tabela de contagem é utilizada para realizar a análise de expressão diferencial. Um total de 212.241.278 pares de sequências de alta qualidade foram utilizados no mapeamento; sendo que 92% destes mapearam no genoma de referência (Tabela 1). Por fim, em média, aproximadamente 80% dos pares de sequências mapearam a genes no genoma de dendê.

Tabela 1. Quantificação das sequências de alta qualidade das amostras de RNA-seq de plantas de dendê controle e estressadas (5 e 12 dias de estresse) contra o genoma de referência de *Elaeis guineensis* (EG5) anotado e disponível no NCBI

Nome da Amostra	Leituras de Entrada Total de “Pair End Sequences”	Leituras Alinhadas					
		A Genes	Não Genes	Ambíguas	Alinhamentos		
					Múltiplos	Baixa Qualidade	Inexistentes
Controle_R1	29.692.333	25.692.842 / 86.53%	1.799.259 / 6.06%	458.227 / 1.54%	1.740.740 / 5.86%	0	1.265 / 0.00%
Controle_R2	30.515.836	26.675.395 / 87.41%	1.771.330 / 5.80%	471.103 / 1.54%	1.596.438 / 5.23%	0	1.570 / 0.01%
Controle_R3	22.045.218	15.956.976 / 72.38%	1.849.181 / 8.39%	270.187 / 1.23%	3.968.049 / 18.00%	0	825 / 0.00%
Estressada_5dias_R1	24.580.201	16.132.336 / 65.63%	3.417.878 / 13.91%	256.196 / 1.04%	4.772.448 / 19.42%	0	1.343 / 0.01%
Estressada_5dias_R2	23.155.078	17.626.321 / 76.12%	2.419.865 / 10.45%	259.472 / 1.12%	2.848.583 / 12.30%	0	837 / 0.00%
Estressada_5dias_R3	23.419.547	15.219.468 / 64.99%	3.561.579 / 15.21%	229.205 / 0.98%	4.408.032 / 18.82%	0	1.263 / 0.01%
Estressada_12dias_R1	21.622.164	15.579.276 / 72.05%	2.188.717 / 10.12%	216.528 / 1.00%	3.636.547 / 16.82%	0	1.096 / 0.01%
Estressada_12dias_R2	23.732.714	14.771.292 / 62.24%	3.367.254 / 14.19%	211.131 / 0.89%	5.381.722 / 22.68%	0	1.315 / 0.01%
Estressada_12dias_R3	23.608.152	18.499.033 / 78.36%	2.027.638 / 8.59%	238.601 / 1.01%	2.842.049 / 12.04%	0	831 / 0.00%

Legenda: Recurso: soma de todas as leituras que foram atribuídas a qualquer gene; Sem recurso: leituras que não puderam ser atribuídas a nenhum gene; Ambíguo: lê as sequências que foram atribuídas a mais de um recurso. Essas leituras não são contadas para nenhum recurso; Alinhamento múltiplo: lê com mais de um alinhamento relatado; Baixa qualidade de alinhamento: leituras ignoradas devido ao filtro "Qualidade mínima de mapeamento". Inexistente: lê o arquivo SAM / BAM sem alinhamento. Fonte: Da autora (2020).

De acordo com as análises utilizando a plataforma de bioinformática OmicsBox (OmicsBox – Bioinformatics Made Easy, BioBam Bioinformatics, March 3, 2019, <https://www.biobam.com/omicsbox>), na sua versão 1.2.4., módulo “transcriptomics”, utilizando o “Reference Based Transcriptomics Workflow” ajustado, onde foi comparado o perfil de expressão diferencial entre Controle vs Estressado 5 dias e Controle vs Estressado 12

dias. Foram encontrados 35 genes diferencialmente expressos ($FDR \leq 0.01$ e $\log FC \geq 5$) quando comparadas as amostras controle com amostras estressadas por 5 dias; e 11 genes quando comparadas as amostras controle com amostras estressadas por 12 dias (Figura 1). Sete genes estavam presentes nos dois grupos de genes diferencialmente expressos encontrados, todos apresentando um aumento na expressão entre 37 e 84X.

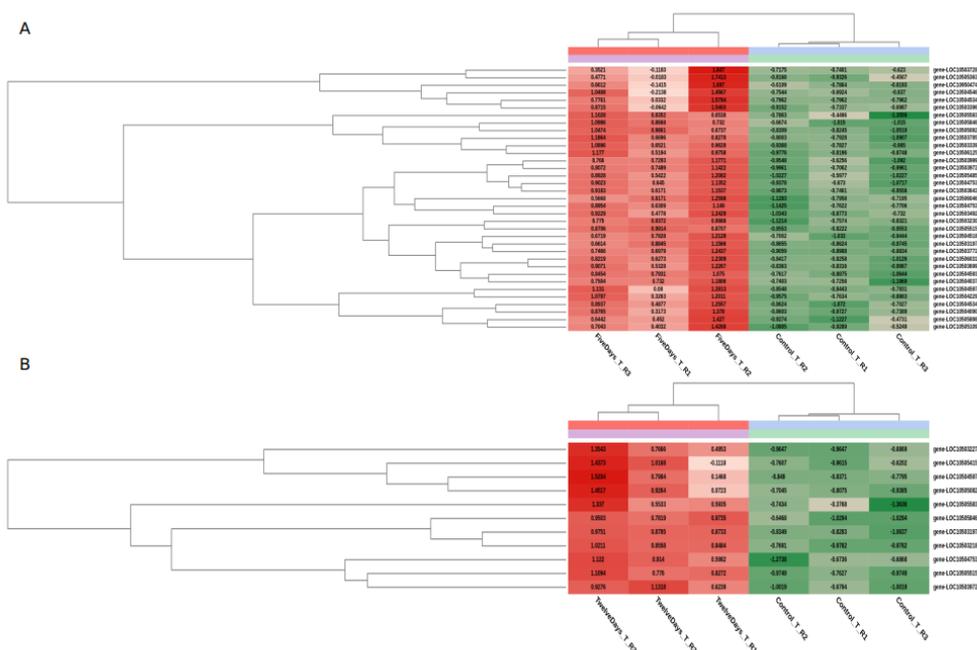


Figura 1. Mapa de calor dos genes diferencialmente expressos (up-regulated) entre amostras de folha de dendê controle e amostras de folha de dendê estressadas com 5 (A) e com 12 dias (B)

Estes 7 genes responsivos a estresse salino foram submetidos à anotação estrutural e funcional, cabendo destacar: a) um gene que expressa a proteína tiofulfato sulfurtransferase; b) um que expressa uma proteína de transferência lipídica não específica; c) um que codifica uma endocitinase ácida; d) um que expressa uma proteína estrutural da parede celular rica em glicina; e) um que codifica um fator de transcrição; f) um que codifica um provável transportador de aminoácido no vacúolo; e g) um gene que codifica uma anidrase carbônica (Tabela 2).

Tabela 2. Genes *up-regulated* selecionados mediante uso dos critérios $FDR \leq 0.01$ e $\log FC \geq 5$, e que estavam diferencialmente expressos aos cinco e aos 12 dias de estresse, quando comparado com o controle

Símbolo do Gene	Descrição do gene	Cromossomo / Scaffold	# de exons	Tipo do gene	Sentido
LOC105055832	Probable vacuolar amino acid transporter YPQ1	13	12	Codificação de proteínas	Anti-sense
LOC105055156	Thiosulfate sulfurtransferase 16. chloroplastic	12	7	Codificação de proteínas	Anti-sense
LOC105045973	Non-specific lipid-transfer protein 1	5	2	Codificação de proteínas	Anti-sense
LOC105047536	Acidic endochitinase	6	1	Codificação de proteínas	Sense

LOC105058460	Transcription factor bHLH94	15	3	Codificação de proteínas	Anti-sense
LOC105039727	Glycine-rich cell wall structural protein 2	2	1	Codificação de proteínas	Anti-sense
LOC105031978	Alpha carbonic anhydrase 7	NW_011550938.1	6	Codificação de proteínas	Sense

As sulfúriltransferases constituem uma família de proteínas grande e complexa presentes principalmente nos cloroplastos (MOSELER et al., 2019). O estudo de Gangadhar et al. (2016) com análise da expressão do gene *StnsLTP1* mostrou diferenças na sua expressão sob estresse por calor, déficit hídrico e salino. Linhagens transgênicas de batata que superexpressam o gene *StnsLTP1* exibiram integridade da membrana celular aprimorada sob estresse. Khalili et al. (2018) desenvolveu linhagens transgênicas de trigo superexpressando endocitina ácida, que está relacionada à defesa contra o estresse oxidativo. Evidências recentes apoiam um modelo no qual certos aspectos da própria parede podem atuar como sinais reguladores do crescimento (SRIVASTAVA, 2019). As condições de estresse abiótico e biótico fornecem novos contextos para o estudo da integridade da parede celular, mas ainda assim, a natureza e as consequências das rupturas da parede devido a vários estressores requerem uma investigação mais aprofundada (SRIVASTAVA, 2019). Por fim, as anidrases carbônicas catalizam a hidratação reversível do dióxido de carbono, gerando prótons e bicarbonato, e provavelmente estão envolvidas na transmissão de sinais para ativar cascata de genes resposta protetora a estresses (RUDENKO et al., 2020)

CONCLUSÕES

A estratégia de prospecção de genes responsivos a salinidade utilizada permitiu a seleção de 7 genes com expressão aumentadas entre 37 e 84X sob estresse salino. Entre estes genes se encontram alguns que expressam proteínas já relatadas como participantes do processo de aumento da tolerância a salinidade.

AGRADECIMENTOS:

À FINEP e à CAPES.

REFERÊNCIAS BIBLIOGRÁFICAS

BORSAI, O.; AL HASSAN, M.; BOSCAIU, M.; SESTRAS, R. E.; VICENTE, O. **The EuroBiotech Journal**, v. 2, n. 2, p.104-113, 2018.

COSTA-SILVA, J.; DOMINGUES, D.; LOPES, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. **PLoS ONE**, v. 12, n. 12, e0190152, p. 1-18, 2017.

GANGADHAR, B. H. et al. Enhanced tolerance of transgenic potato plants over-expressing non-specific lipid transfer protein-1 (StnsLTP1) against multiple abiotic stresses. **Frontiers in Plant Science**, v. 7, p. 1228, 2016.

JIMÉNEZ-JACINTO, V.; SANCHEZ-FLORES, A.; VEGA-ALVARADO, L. Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis. **Front. Genet.**, v. 10, n. 279, p. 1-16, 2019.

KHALILI, M.; NAGHAVI, M. R.; YOUSEFZADEH, S. Protein pattern analysis in tolerant and susceptible wheat cultivars under salinity stress conditions. **Acta agriculturae Slovenica**, v. 111, n. 3, p. 545-558, 2018.

LOWE, R.; SHIRLEY, N.; BLEACKLEY, M.; DOLAN, S.; SHAFEE, T. Transcriptomics technologies. **PLoS Comput Biol**, v. 13, n. 5, e1005457, 2017.

MOSELER, A. et al. Novel insights into the diversity of the sulfurtransferase family in photosynthetic organisms with emphasis on oak. **New Phytologist**, v. 226, n. 4, 967-977, 2019.

RUDENKO, N. N.; BORISOVA-MUBARAKSHINA, M. M.; IGNATOVA, L. K.; FEDORCHUK, T. P.; NADEEVA-ZHURIKOVA, E. M., IVANOV, B. N. **Role of Plant Carbonic Anhydrases under Stress Conditions**. 26p., 2020.

SHAHID S. A.; ZAMAN M.; HENG L. 2018. Soil Salinity: Historical Perspectives and a World Overview of the Problem. In: **Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques** (<https://doi.org/10.1007/978-3-319-96190-3>).

SINGH, R. *et al.* Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. **Nature**, v. 500, n. 7462, p. 335-339, 2013.

SRIVASTAVA, N. Molecular and Biotechnological Tools in Developing Abiotic Stress Tolerance in Wheat. In: HASANUZZAMAN, M.; NAHAR, K.; HOSSAIN, Md. A. **Wheat Production in Changing Environments**. Singapore: Springer, 2019. p. 283-341.

VIEIRA, L. R.; SILVA, V. N. B.; CASARI, R. A. das C. N.; CARMONA, P. A. O.; SOUSA, C. A. F. de; SOUZA JUNIOR, M. T. Morphophysiological responses of young oil palm plants to salinity stress. **Pesquisa Agropecuária Brasileira**, v. 55, e01835, 2020.

WANG, J.; MENG, X.; DOBROVOLSKAYA, O. B.; ORLOV, Y. L.; CHEN, M. Non-coding RNAs and Their Roles in Stress Response in Plants. **Genomics, Proteomics & Bioinformatics**, v. 15, n. 5, p. 301-312, 2017.