VALIDAÇÃO DE RETROTRANSPOSONS ATIVOS EM CAFÉ- EFEITOS DE ESTRESSE ABIÓTICO E BAIXA CAFEÍNA NA EXPRESSÃO GÊNICA[1]

Paula de Sousa Guimarães[2]; Juliana Camargo Martinati Schenk[3]; Poliana Fernanda Giachetto[4]; Lilian Padilha[5]; Maria Bernadete Silvarolla[6]; Mirian Perez Maluf[7]

1 Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café
2 Bolsista Consórcio Pesquisa Café, DSc, psguim@yahoo.com.br
3 Bolsista INCT/CNPq, DSc, Instituto Agronômico de Campinas, Campinas-SP, julianamartinati@gmail.com
4 Pesquisadora, DSc, Embrapa Informática, Campinas-SP, poliana.giachetto@embrapa.br
5 Pesquisadora, DSc, Embrapa Café, Brasília-DF, lilian.padilha@embrapa.br
6 Pesquisadora, MSc, Instituto Agronômico de Campinas, Campinas-SP, bernadet@iac.sp.gov.br
7 Pesquisadora, PhD, Embrapa Café, Brasília-DF, mirian.maluf@embrapa.br

RESUMO: A identificação de perfis de expressão gênica diferenciais e redes metabólicas é uma ferramenta valiosa para a caracterização genética de caracteres agronômicos. Neste estudo, utilizamos análises de expressão em larga-escala para identificar processos biológicos alterados em plantas com teores reduzidos de cafeína. A expressão diferencial de genes selecionados observada *in silico* foi então validada em experimentos *in vivo*. O primeiro passo foi o sequenciamento em larga-escala de RNA extraído de tecidos jovens e em desenvolvimento, de plantas sem cafeína e de plantas com concentrações normais do composto. As sequências obtidas foram agrupadas, a partir de análises *in silico*, para identificação de genes com expressão diferencial entre os tratamentos e de redes metabólicas associadas aos teores de cafeína. Para a validação das análises *in silico*, selecionamos 3 genes de famílias de retrotransposons para análises de qRT-PCR em gemas e folhas coletadas de plantas normais (MN) e sem cafeína (AC), tanto em condições normais quanto durante déficit hídrico. O perfil de expressão dos genes estabelecido por qRT-PCR foi distinto do observado nas análises de RNAseq. Além disso, os resultados sugerem que plantas AC possuem uma maior suscetibilidade ao estresse hídrico, uma vez que a expressão de retrotransposons é mais alta em folhas e gemas de plantas AC do que em MN.

PALAVRAS-CHAVE: seleção-assistida, RNAseq, melhoramento molecular.

VALIDATION OF ACTIVE RETROTRANSPOSONS IN COFFEE- EFFECTS OF ABIOTIC STRESS AND LOW CAFFEINE CONTENT ON GENE EXPRESSION

ABSTRACT: Differential gene expression profiles and metabolic networks are valuable tools for the genetic characterization of agronomic traits. In this study, we used large-scale expression analyses to identify altered biological processes in plants with reduced levels of caffeine. Then, the differential expression of selected genes observed *in silico* was validated by *in vivo* experiments. The first step was the large-scale sequencing of RNA extracted from young and developing tissues, of caffeine-free plants (AC) and of plants with normal concentrations of the compound (MN). The sequences obtained were grouped and analyzed *in silico* for identification of genes with differential expression between treatments, and establishment of metabolic networks associated with levels of caffeine. For validation of RNAseq, we selected 3 retrotransposons genes for qRT-PCR analyses in bud and leaves collected from regular and caffeine-free plants, both in normal conditions or during water deficit. The profile of gene expression by qRT-PCR was different to that observed with RNAseq. In addition, the results suggest that AC plants have a higher susceptibility to water stress, once retrotransposons are more expressed in AC leaves and buds than in MN.

KEY WORDS: assisted-selection, RNAseq, molecular breeding.

INTRODUÇÃO

Análises genômicas em larga-escala são uma ferramenta valiosa para a identificação de redes metabólicas e processos biológicos envolvidos com características de interesse agronômico. A identificação de grupos de genes de diferentes funções biológicas e com expressão diferencial associada a um caractere é essencial para implementação de estratégias de genômica associativa ampla por programas de melhoramento. Além disso, a validação da expressão de genes específicos promove um maior conhecimento da base genética de características de herança complexa. Assim, análises genômicas de larga-escala, como RNAseq e GBS, vem sendo utilizadas para uma ampla caracterização de genótipos, processos biológicos e controle genético de características de interesse agronômico. Neste estudo, utilizamos a estratégia de RNAseq para avaliar o papel da cafeína em processos fisiológicos de cafeeiros. Estudos anteriores realizados em cafés com teores de cafeína naturalmente reduzidos, identificados pelo Programa de Melhoramento do IAC, sugerem que a ausência de cafeína pode afetar aspectos botânicos e fisiológicos da planta (Silvarolla et al, 2004; Favoretto et al, 2017). Portanto, buscamos identificar processos biológicos afetados durante a organôgenese de folhas e frutos em plantas normais e sem cafeína. As análises de RNAseq indicaram um número reduzido, porém significativo, de genes com expressão diferencial (GDE) entre os genótipos, incluindo fatores de transcrição reguladores de

diferenciação celular e retrotransposons. Para validação, foi avaliada a expressão de sequências homólogas a retrotransposons em acessos da cultivar Mundo Novo (MN) e da variedade AC (Programa de Melhoramento IAC), em condições de plena irrigação e de déficit hídrico. Uma vez que a ativação de transposons em condições de estresse abiótico é amplamente descrita na literatura (Mirouze & Paszkowski, 2011), o objetivo desta análise foi validar a expressão diferencial observada em diversas condições experimentais.

MATERIAL E MÉTODOS

Sequenciamento de transcriptoma (RNAseq) - Amostras de tecidos vegetais, tais como folhas jovens, folhas adultas, gemas apicais e ramos com gemas foram coletadas de 48 plantas adultas com teores reduzidos de cafeína (progênies AC) e normais (Cultivar Mundo Novo). O RNA total de plantas selecionadas foi extraído individualmente usando o kit de extração RNeasy Plant Mini Kit da QUIAGEN, seguindo o protocolo especificado pelo fabricante. A purificação do RNA foi realizada através de membranas QIAshredder fornecidas pelo fabricante. O RNA total obtido foi armazenado em freezer a -20ºC. A seleção das amostras para sequenciamento de alta cobertura (RNAseq) foi baseada na qualidade e integridade do RNA total extraído. Após a seleção, as amostras foram diluídas em 55 µl a uma concentração de 50 ng/µl. As amostras foram unidas duas a duas, baseando-se no teor de cafeína (Normal/Baixo) e no tipo de tecido vegetal que foi utilizado para extração do RNA total. Após a união, alíquotas de 55 µl das amostras foram enviadas em duplicata para o sequenciamento do RNAseq. As etapas de conversão de RNA em cDNA, precipitação e purificação do cDNA, quantificação do cDNA através de espectrofotometria, construção das bibliotecas e sequenciamento utilizando a tecnologia Illumina HiSeq2500 foram desenvolvidas pelo laboratório BIODOME.

Análises *in silico* - As *reads* analisadas foram geradas a partir do sequenciamento *paired end* (2x125pb) de 6 bibliotecas. Os transcritos foram obtidos por meio de montagem de novo utilizando-se o programa Trinity e os genes diferencialmente expressos (GDE) entre os genótipos MN e AC em folhas e ramos foram identificados utilizando-se o pacote do R/Bioconductor EdgeR (Robertson et al., 2010), executado no ambiente R versão 3.1.0 "Spring Dance" (http://www.r-project.org/). A anotação funcional, categorização em termos de Gene Ontology (GO) e o mapeamento nas vias metabólicas do KEGG foram realizados utilizando-se o programa Blast2GO. Para a seleção de *primers* gene-específicos, as sequências Fasta correspondentes aos genes de retrotransposons foram comparadas com as sequências de ESTs de banco de dados públicos (NCBI), utilizando-se, para isso, o algoritmo tBLASTx. Para classificar a parte funcional da proteína, bem como sua matriz de leitura, foi utilizada a ferramenta Open Reading Frame Finder (ORF FINDER) e ainda se realizou a busca dos domínios conservados das proteínas, a fim de definir as regiões para anelamento dos *primers*.

Análises de expressão gênica - Para extração do RNA foram coletadas folhas e gemas de plantas AC e MN mantidas em campo experimental sem irrigação, na Fazenda Santa Elisa do IAC em Campinas. As coletas foram realizadas em condições de irrigação natural plena (Janeiro/18) e após estiagem de 45 dias (Setembro/17; dados Portal Agritempo). O RNA total foi extraído pelo método citado por Chang et al. (1993). A quantificação do RNA foi realizada medindo-se a absorbância a 260nm, 280nm e 230 nm. As amostras de RNA total foram tratadas com DNAse I (PromegaTM). A partir dos RNAs tratados com DNAse I foi sintetizada a primeira fita de cDNA utilizando o kit SuperScript™ II Reverse Transcriptase (Thermo). Em seguida, as amostras foram armazenadas a -20°C. A quantificação da expressão de 3 genes selecionados foi realizada por qRT-PCR, em equipamento ABI7300. As reações foram realizadas em volume final de 15 µL contendo 400 ng de cDNA, 200mM de cada primer e 1X de SYBR Green master mix (Promega). Todas as amostras foram amplificadas em triplicatas. Foram utilizadas também amostras sem adição de cDNA para detectar qualquer sinal de contaminação (controle negativo). Para confirmar a presença de amplicons únicos, os produtos do PCR foram submetidos à curva de dissociação com temperatura variando entre 60ºC e 95ºC. O gene constitutivo GAPDH foi usado com normalizador (controle endógeno). Os valores obtidos para folhas irrigadas da cultivar MN foram utilizados como calibradores para cálculo da concentração relativa de todos os tratamentos (Iskandar et al, 2004).

RESULTADOS E DISCUSSÃO

Os resultados obtidos com a montagem *de novo* dos transcriptomas resultou na identificação de 65.536 isoformas. Apesar da complexidade desta abordagem, particularmente quando o genoma é grande e o material analisado possui sequências repetitivas (Metzker, 2010), os dados obtidos aqui indicam que a montagem é representativa do genoma do café. As análises quantitativas das isoformas identificaram um total de 83 e 88 GDE entre os genótipos MN e AC para folhas e ramos, respectivamente, sendo que todos foram expressos exclusivamente em um ou outro genótipo. O total de genes com expressão diferencial não é alto, o que indica tanto uma reduzida variabilidade genética entre os genótipos, quanto um efeito limitado da ausência de cafeína na expressão de genes nos tecidos avaliados. A categorização funcional das sequências evidenciou os processos biológicos, incluindo Processos metabólicos primários e Processos metabólicos de substâncias orgânicas. Já as funções moleculares Ligação a compostos heterocíclicos e Ligação a compostos cíclicos orgânicos foram as mais frequentes entre os GDE nas folhas e ramos. O mapeamento dos GDE nas vias metabólicas do KEGG identificou as vias Metabolismo de tirosina e Biossíntese de isoquinolina alcalóide como as mais representativas em folhas, entre os genótipos MN e AC. Nos ramos, a via que apresentou o maior número de GDE foi a Biossíntese de antibióticos.

Um total de 3 genes foram selecionados a partir das análises de RNAseq, com expressão diferencial entre tecidos e órgãos AC e MN, para validação utilizando-se PCR em tempo real. Os genes estão relacionados com elementos de transposição. O *contig 1* tem homologia com a família de retrotransposon non-LTR, *long terminal repeat retrotransposon*, e os *contigs 2* e *3* com a família de transposons *Zinc finger BED domain-containing protein*. Para as análises utilizamos cDNAs sintetizados a partir de RNA total extraído de folhas e gemas, coletadas em plantas AC e MN, mantidas em condições normais e de estresse hídrico. Os resultados destas análises estão apresentados na Figura 1.

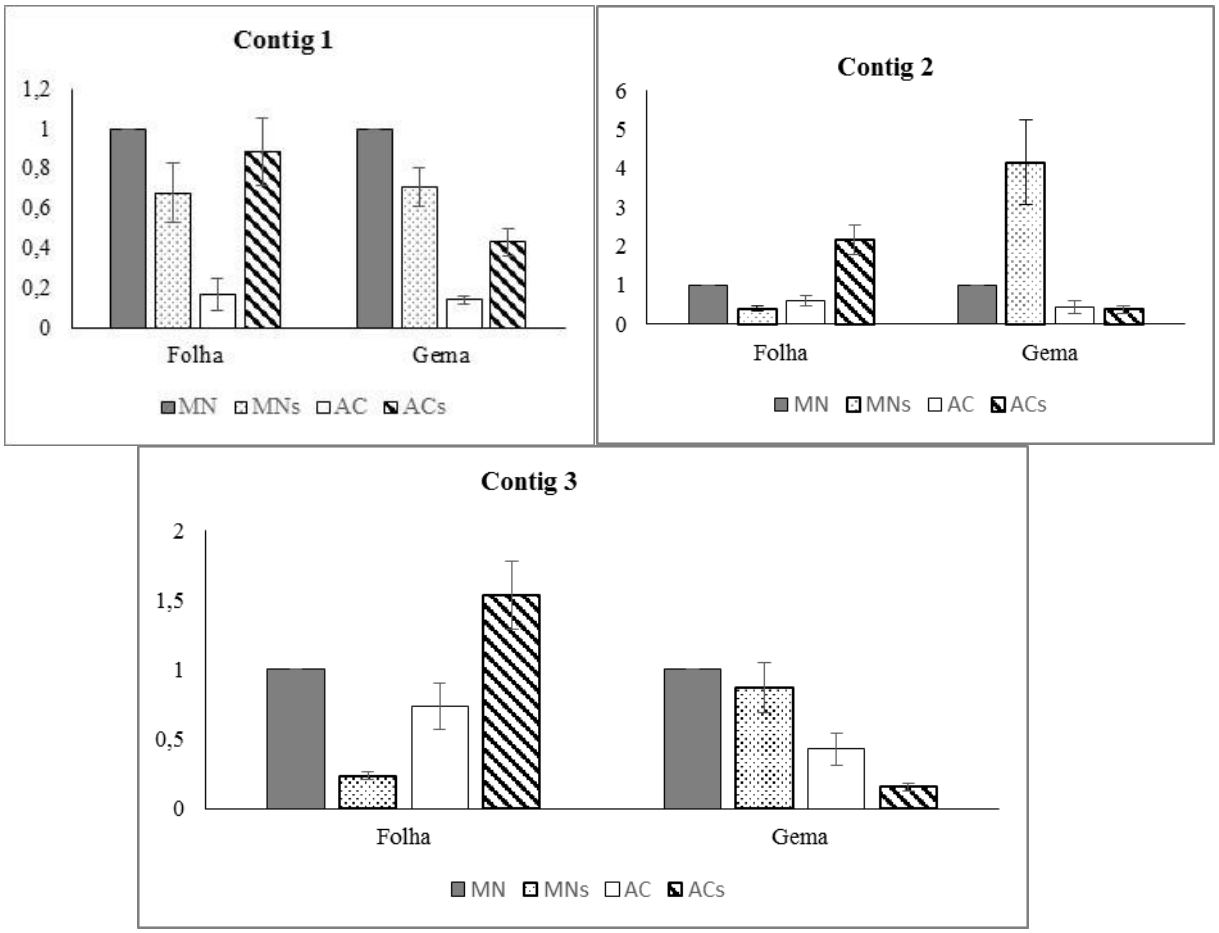

Figura 1. Comparação da expressão de genes associados a retrotransposons entre plantas com teores normais de cafeína (MN) e sem cafeína (AC), em condições normais e de déficit hídrico (MNs e ACs)

Para os 3 genes avaliados os valores de expressão diferencial observados por RNAseq em folhas e gemas não foram validados, uma vez que estes genes se expressam somente em tecidos de plantas AC. Pelas análises de qRT-PCR os valores de expressão quantificados por qRT-PCR foram maiores em tecidos MN, indicando uma inversão para o perfil de expressão. O *contig 1* apresenta essa inversão mais evidente. A validação de perfis de expressão através de análises *in vivo* é essencial para corrigir desvios resultantes da conversão de dados brutos de sequenciamento em valores relativos e normalizados. Outro aspecto que contribui para a ocorrência de desvios nos perfis de expressão é a estratégia normalmente utilizada, onde bibliotecas de cDNA oriundas de diferentes tecidos são misturadas em uma única amostra, para otimizar a reação de sequenciamento e formar um pool único de sequências. Assim, pode haver alterações na representatividade de sequências específicas neste *pool*, em razão de diferenças na quantidade inicial de mRNA de cada tecido amostrado.

Nas comparações entre tecidos irrigados e em condições de estresse hídrico, as análises de expressão indicam uma resposta diferencial entre genótipos e tecidos. Para os 3 genes avaliados, o perfil observado em tecidos de MN indica que o estresse hídrico leva a uma repressão na expressão dos retrotransposons. Já em folhas de AC observa-se uma expressiva ativação destes genes em resposta ao estresse hídrico, e em gemas estressadas a expressão é ativada no *contig 1*, porém é reprimida no *contig 3*. Essas alterações observadas no perfil de expressão sugerem que plantas do genótipo AC tem uma resposta mais acentuada ao estresse hídrico. No entanto, quanto desta resposta pode ser decorrente da ausência de cafeína, uma vez que tecidos de MN com teores normais de cafeína apresentam repressão destes genes, necessita de mais análises incluindo genes relacionados com síntese e transporte de cafeína. Elementos de transposição são ativados em condições de estresse biótico, e também são associados ao aparecimento de padrões

alterados de desenvolvimento (Mirouze & Paszkowski, 2011). Neste estudo, selecionamos estes genes para monitorar se a presença de elementos de transposição pode estar associada com alterações na arquitetura de plantas sem cafeína. As análises aqui não indicam essa ação, uma vez que em condições normais de irrigação plantas AC tem uma expressão relativa dos genes relacionados a transposons menor do que plantas normais.

CONCLUSÃO

1. A ausência de cafeína não resulta em alterações na expressão de muitos genes, e também não afeta processos metabólicos significativos.
2. A ativação de elementos de transposição não está relacionada com alterações na arquitetura de plantas com teores reduzidos de cafeína.
3. A regulação do padrão de expressão dos elementos de transposição pelo estresse hídrico difere entre os tecidos de um mesmo genótipo.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro do Consórcio Pesquisa Café e do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

REFERÊNCIAS BIBLIOGRÁFICAS

CHANG, S.; PURYERAR, J.; CAIRNEY, J. A single and efficient method for isolating RNA from pine trees. Plant Molecular Biology. v.11, p.113-116, 1993.

FAVORETTO, P.; SILVA, C. C.; TAVARES, A. G.; GIATTI, G.; MORAES, P. F.; LOBATO, M. T. V.; SILVAROLLA, M. B.; GUERREIRO-FILHO, O.; MALUF, M. P. Assisted-selection of naturally caffeine-free coffee cultivars characterization of SNPs from a methyltransferase gene. Molecular Breeding, v. 37, p. 31, 2017.

ISKANDAR, H.M.; SIMPSON, R.S.; CASU, R.E.; BONNETT, G.D.; MACLEAN, D.J.; MANNERS, J.M. Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. Plant Mol Biol Report 22, p. 325–337, 2004.

METZEKER, M. Sequencing technologies —the next generation. Nature Reviews Genetics 11, p. 31- 46, 2010.

MIROUZE, M.; PASZKOWSKI, J. Epigenetic contribution to stress adaptation in plants. Curr. Opin. Plant Biol. 14, p. 267-274, 2001.

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

SILVAROLLA, M.B.; MAZZAFERA, P.; FAZUOLI L.C. A naturally decaffeinated arabica coffee. Nature, 429, p.826–826, 2004.