

EXPRESSÃO GÊNICA EM LARGA ESCALA DE RAIZES DE *Coffea arabica* L.¹

Tiago Benedito dos Santos²; Juliana Costa Silva²; João Danillo Moura Soares³; Viviane Yumi Baba⁴; Luiz Filipe Protasio Pereira⁵; Luiz Gonzaga Esteves Vieira⁶; Douglas Silva Domingues⁷

¹Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

²Bolsista Consórcio Pesquisa Café, BS, tiagobio02@yahoo.com.br, julianacostasilvati@gmail.com

³Doutoranda em Agronomia (UEL), Londrina-PR, vybaba15@gmail.com.br

⁴Mestrando em Genética e Biologia Molecular (UEL), Londrina-PR, j.danillo@yahoo.com.br

⁵Pesquisador, PhD, Embrapa Café, Brasília-DF, filipe.pereira@embrapa.br

⁶Pesquisador, DSc, Universidade do Oeste Paulista, Presidente Prudente (UNOESTE), São Paulo, SP, luizgonzaga@unoeste.br

⁷Pesquisador, DSc, Instituto de Biociências de Rio Claro, Universidade Estadual Paulista (UNESP), Rio Claro – SP, dougsd3@gmail.com

RESUMO: A utilização de novas tecnologias de sequenciamento em larga escala (454-mRNAseq) é uma frente inovadora de estudos para a biotecnologia do cafeeiro, promovendo a identificação e caracterização de novos componentes genômicos de interesse. Mudanças de cafeeiro foram mantidas sob supressão de fontes nitrogenadas e suas raízes foram utilizadas para construção de bibliotecas de cDNA, que foram sequenciadas utilizando a plataforma GS-FLX *Titanium* da Roche. Foram gerados um total 809.517 *reads*, os quais geraram 34.654 *contigs*. Por meio dos transcritos gerados pelo mRNAseq de raiz espera-se identificar e caracterizar molecularmente os genes envolvidos no transporte e metabolismo de nitrogênio no cafeeiro.

PALAVRAS-CHAVE: Café, nutrição mineral, sequenciamento de nova geração, 454

LARGE-SCALE GENE EXPRESSION ANALYSIS IN *Coffea arabica* L. roots

ABSTRACT: The use of new large-scale sequencing technologies is an innovative strategy in coffee biotechnology, identifying and characterizing genomic components involved in nitrogen (N) starvation. Coffee plants were grown under N starvation and roots were used for construction of cDNA libraries and sequencing using GS-FLX *Titanium* Roche platform. We obtained 809 517 *reads*, assembled in 34,654 *contigs*. Using this data, we expect to identify and characterize genes involved in transport and nitrogen metabolism in coffee plants.

KEYWORDS: Coffee, mineral nutrition, next generation sequencing, 454

INTRODUÇÃO

O desequilíbrio nutricional na cultura do café é um dos principais entraves para a sua produção. Como não existem informações sobre os mecanismos moleculares de absorção de nutrientes em cafeeiro, uma possibilidade para obtenção de plantas de café mais eficientes em absorver e utilizar os nutrientes é estudar os genes que codificam possíveis transportadores de nutrientes minerais. Atualmente, a técnica de RNA-seq tem sido revolucionária na determinação das bases transcricionais de um determinado órgão, tecido ou estado fisiológico, devido sua capacidade de sequenciamento de grande cobertura, que permitem amostragens de virtualmente mais de 90% dos genes transcricionalmente ativos em um dado tecido (Wang et al, 2009). Em plantas, técnicas de sequenciamento de nova geração (NGS) são largamente utilizadas para desenvolvimento de marcadores moleculares, e compreensão de respostas transcricionais a estresses bióticos e abióticos, conforme revisado em Schliesky et al. (2012). Uma das plataformas mais utilizadas em plantas tem sido a Roche/454. Devido a sua grande cobertura de sequenciamento, fica evidente que a utilização desta metodologia pode ajudar na identificação de novos componentes moleculares envolvidos na adaptação do cafeeiro à condição de supressão de N.

Visto que a maioria dos estudos transcricionais no cafeeiro está direcionada a estudos sobre o desenvolvimento de frutos (Budzinski et al, 2011; Cação et al, 2012), caracterização da expressão de famílias gênicas candidatas (dos Santos et al, 2011; Marracini et al, 2012) ou de um conjunto pré-determinado de genes em microarranjo (Bardil et al, 2011), a utilização da técnica de sequenciamento 454 é uma frente inovadora de estudos para esta cultura, promovendo a identificação e caracterização de novos componentes regulados sob supressão de N.

O presente trabalho pretende estabelecer uma nova frente de pesquisa pouco estudada em termos moleculares, sobre as respostas transcricionais frente à adaptação às condições de supressão de N, em uma planta perene de grande interesse no agronegócio nacional, o cafeeiro. Para isso, serão desenvolvidos uma avaliação transcricional em larga escala em raízes (mRNAseq) e posteriormente a caracterização molecular de genes envolvidos no transporte de N.

MATERIAL E MÉTODOS

A metodologia aqui representada segue os procedimentos já utilizados em artigo anterior publicado pelo grupo (de Carvalho et al, 2013), conforme figura 1.

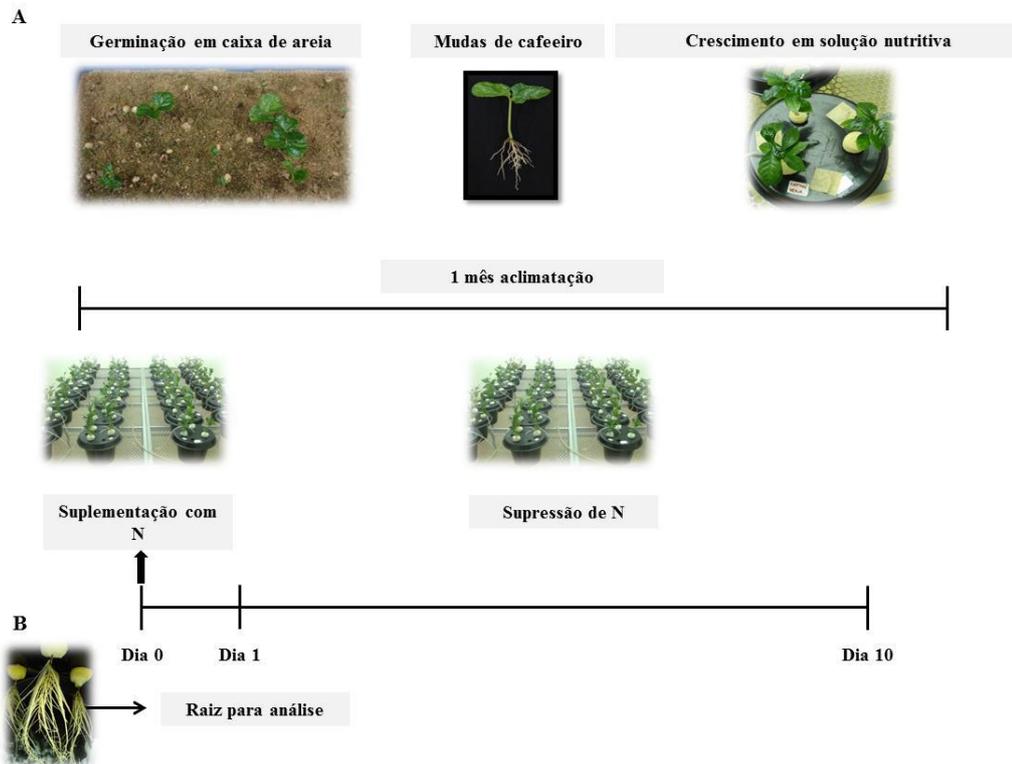


Figura 1. A. Plantas da cultivar IAPAR59 foram germinadas em areia e depois transferidas para solução nutritiva modificada de Clark (1975) por 30 dias para aclimação. A seguir, as plantas foram transferidas para solução nutritiva sem fontes de N, por um período de 10 dias. Foram coletadas raízes antes da troca da solução (controle - dia 0), 24h após a troca da solução (1 dia) e 10 dias após a troca da solução para extração de RNA total e demais análises.

Extração de RNA total, construção das bibliotecas de cDNA e sequenciamento de nova geração – mRNAseq

O RNA total de raiz do tratamento sob supressão de N foi extraído conforme descrito por Chang et al. (1993). A partir de 20 µg de RNA total foram isolados de 500ng mRNA utilizando o kit de purificação DynaBeads mRNA purification kit (Life technologies) conforme as instruções do fabricante. As bibliotecas de cDNA de raiz foram preparadas com 500 ng utilizando a enzima Superscript III (Invitrogen). Adaptadores direcionais modificados para o sequenciado 454 GS FLX (adaptador A e B) foram ligados ao cDNA. As bibliotecas de cDNA de cadeia simples foram amplificadas por PCR e purificados, resultando em bibliotecas dscDNA. Posteriormente, as bibliotecas foram quantificadas utilizando quantificação fluorimétrica Qubit®, e os tamanhos médios dos fragmentos foram determinados utilizando um chip de alta sensibilidade pelo Agilent Bioanalyzer 2100 (Agilent Technologies). As bibliotecas de cDNA ligado ao adaptador foram reunidas em concentrações equimolares, submetido a amplificação clonal à base de emulsão e pirosequenciado sobre uma placa completa com o sistema 454 GS FLX (454 Life Sciences, Roche, Branford, CT, EUA) de acordo com as instruções do fabricante.

Análises de bioinformática

Inicialmente os dados foram filtrados e trimados por critérios de qualidade, para exclusão de seqüências dos adaptadores utilizados no sequenciamento e leituras de baixa qualidade. A montagem e análise dos transcritos foram realizadas através do pipeline Trinity (Haas et al, 2013). Para a anotação funcional dos *unigenes* foi utilizada a plataforma Blast2GO (*Gene Ontology terms*) v.2.6.4 (Conesa & Gotz, 2008), em parâmetros padrão. Além destas ferramentas, os *unigenes* gerados foram analisados por BlastX (Altschul et al, 1997) com (e-value $1e^{-5}$) para alinhamento dos contigs contra os seguintes bancos de dados públicos: o banco de dados não redundantes (NR) do NCBI; o Uniref90 e Uniref100, que contêm conjuntos de seqüências de proteínas providas da plataforma UniProt (Jain et al, 2009), o banco de dados Pfam de famílias de proteínas (Punta et al, 2012), o banco de dados KEGG (<http://www.genome.jp/kegg/>) de vias metabólicas e a versão 6.0 de proteínas do PlantCyc (<http://www.plantcyc.org/>).

RESULTADOS E DISCUSSÃO

Através do sequenciamento de nova geração de cDNA de raízes do cafeeiro submetidas a supressão de N, foram obtidos um conjunto de 34.654 *contigs* (Tabela 1), correspondendo aproximadamente 341 MB. Os *reads* apresentaram em média 456 pb (pares de bases) e *contigs* tiveram o tamanho médio de 966 pb (Tabela 1).

Tabela 1. Estatística dos dados gerados através do transcriptoma de raiz de *C. arabica*.

Contig number	34.654
Mean contig length	966
Maximum contig length	8.514
Minimum contig length	379
N50 length	1.314
GC% content	43

Análises preliminares de bioinformática indicaram que a maioria dos contigs obtidos são representados por genes que possuem apenas uma isoforma; 9,72% apresentam 2 isoformas; 2,98% apresentam 3 isoformas e 2,32% são representados acima de 4 isoformas, respectivamente (Figura 2).

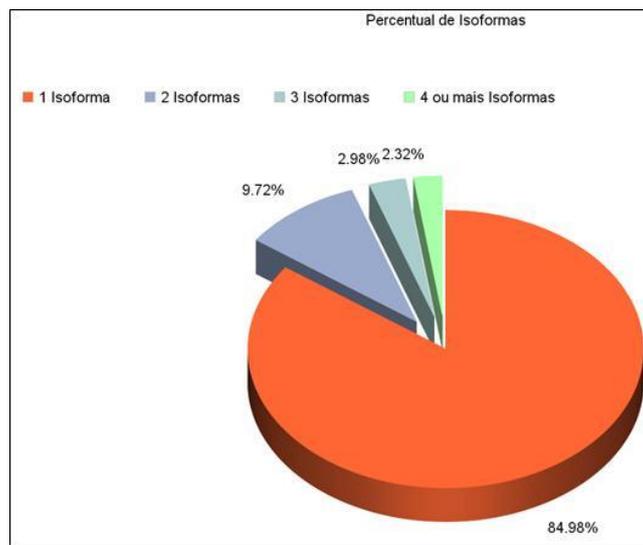


Figura 2. Abundância do número de isoformas no transcriptoma de raiz de *C. arabica*. A porcentagem representa a variação do número de isoformas para um único gene.

Foram realizadas combinações entre as bibliotecas de raízes (0x1; 1x10; 0;10 dias sem N), a fim de se identificar genes diferencialmente expressos, que obtiveram FDR (False Discovery Rate) < 0,1, de acordo com a ferramenta edgeR. Um total de 62 genes são diferencialmente expressos. A maioria deles foi observada na comparação 0x10 (31 genes). Dez genes foram diferencialmente expressos entre 0 e 1 dia sem fontes de N e na combinação 1x10 foram identificados 21 genes diferencialmente expressos.

Diante deste catálogo de genes identificados no transcriptoma de raiz de *C. arabica*, pode-se agora desenhar experimentos que permitam utilizar estes dados na identificação de fatores moleculares envolvidos na absorção e transporte de N em café arábica.

CONCLUSÕES

- Foi possível avaliar em larga escala mais de 34.000 genes transcricionalmente ativos em raiz de *C. arabica*.
- Genes diferencialmente expressos poderão auxiliar no entendimento de fatores envolvidos na absorção e transporte de N em *C. arabica*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W., LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 1997, 25, 3389–3402.
- BARDIL, A., DE ALMEIDA, J. D., COMBES, M. C., LASHERMES, P., BERTRAND, B. Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. *New Phytologist*, 2011, v. 192, n. 3, p. 760-774.

- BUDZINSKI, I. G. F., SANTOS, T. B., SERA, T., POT, D., VIEIRA, L. G. E., PEREIRA, L. F. P. Expression patterns of three α -expansin isoforms in *Coffea arabica* during fruit development. *Plant Biology*, 2011, 13(3), 462-471.
- CAÇÃO, S. M. B., LEITE, T. F., BUDZINSKI, I. G. F., DOS SANTOS, T. B., SCHOLZ, M. B. S., CARPENTIERI-PIPOLO, V., DOMINGUES, D. S., VIEIRA, L. G. E., PEREIRA, L. F. P. Gene expression and enzymatic activity of pectin methylesterase during fruit development and ripening in *Coffea arabica* L. *Genetics and Molecular Research*, 2012, 11(3), 3186-3197.
- CONESA, A., STEFAN, G. "Blast2GO: A comprehensive suite for functional analysis in plant genomics." *International journal of plant genomics*, 2008.
- CHANG, S.; PURYEAR, J.; CAIRNEY, J. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology*, v.11, p.113–116, 1993.
- de CARVALHO, K., BESPALHOK FILHO, J. C., dos SANTOS, T. B., DE SOUZA, S. G. H., VIEIRA, L. G. E., PEREIRA, L. F. P., & DOMINGUES, D. S. Nitrogen starvation, salt and heat stress in coffee (*Coffea arabica* L.): identification and validation of new genes for qPCR normalization. *Molecular biotechnology*, 2013, 53(3), 315-325.
- dos SANTOS TB, BUDZINSKI IGF, MARUR CJ, PETKOWICZ CLO, PEREIRA LFP, VIEIRA LGE. Expression of three galactinol synthase isoforms in *Coffea arabica* L. and accumulation of raffinose and stachyose in response to abiotic stresses. *Plant Physiology and Biochemistry* 49:441-448, 2011.
- HAAS, B. J., PAPANICOLAOU, A., YASSOUR, M., GRABHERR, M., BLOOD, P. D., BOWDEN, J., et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 2013, 8(8), 1494-1512.
- MARRACCINI P, VINECKY F, ALVES GS, RAMOS HJ, ELBELT S, VIEIRA NG, CARNEIRO FA, SUJII PS, ALEKCEVETCH JC, SILVA VA, DAMATTA FM, FERRAO MA, LEROY T, POT D, VIEIRA LG, DA SILVA FR, ANDRADE AC. Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *J Exp Bot*. Apr 17, on line, 2012.
- PUNTA, M., COGILL, P. C., EBERHARDT, R. Y., MISTRY, J., TATE, J., BOURSNEILL, C., et al. RD Finn The Pfam protein families database. *Nucleic Acids Research*, 2012.
- SCHLIESKY, S., GOWIK, U., WEBER, A. P., BRÄUTIGAM, A. RNA-seq assembly—are we there yet?. *Frontiers in plant science*, 2012, v. 3.
- WANG, Z. *et al.* RNA-seq: A revolutionary tool for transcriptomics. *Nature*, v.10, p.57-63, 2009.