

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES ÓRFÃOS (“NO HITS”) DE CAFÉ (*COFFEA CANEPHORA*), ENVOLVIDOS NA RESPOSTA À SECA¹

Karoline Estefani Duarte², Natália Gomes Vieira³, Polyana Kelly Martins⁴, Ana Paula Ribeiro⁵, Bárbara Andrade D. B. da Cunha⁶, Hugo Bruno Correa Molinari⁷, Adilson Kenji Kobayashi⁸, Pierre Marraccini⁹, Alan Carvalho Andrade¹⁰

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café, FINEP, Programa CAPES-EMBRAPA e INCT-Café (CNPq/FAPEMIG).

² Bolsista CAPES, Mestranda, karollduarte31@gmail.com

³ Bolsista Consórcio Pesquisa Café, MS, ngvieiral@gmail.com

⁴ Bolsista CAPES, Pós-doutoranda, polykmartins@gmail.com

⁵ Bolsista CAPES, Doutoranda, anapaulabiotec@gmail.com

⁶ Analista, MSc, Embrapa Agroenergia, Brasília-DF, barbara.dias@embrapa.br

⁷ Pesquisador, PhD, Embrapa Agroenergia, Brasília-DF, hugo.molinari@embrapa.br

⁸ Pesquisador, PhD, Embrapa Agroenergia, Brasília-DF, adilson.kobayashi@embrapa.br

⁹ Pesquisador, PhD, CIRAD UMR AGAP, Montpellier – FR, marraccini@cirad.fr

¹⁰ Pesquisador, PhD, Embrapa Recursos Genéticos e Biotecnologia (LGM), Brasília-DF, alan.andrade@embrapa.br

RESUMO: O café é uma planta perene, sendo considerada uma das *commodities* agrícolas mais importantes do mundo. Por consequência, visando-se o estabelecimento de ferramentas de auxílio para se acelerar o melhoramento genético desta espécie, avanços significativos em genômica do cafeeiro têm ocorrido nos últimos anos. Como exemplo, pode-se citar a recente conclusão do sequenciamento do genoma completo de *Coffea canephora*, que servirá de referência para utilização em trabalhos avançados de genética molecular, aplicados diretamente ao melhoramento genético desta espécie, tais como o estabelecimento de programas de seleção genômica ampla (SGA) em cafeeiro. Análises recentes de bioinformática dos genomas completos de plantas indicam que cerca de 20-30% do total de genes de cada genoma são inéditos e específicos de cada espécie. Ou seja, essas sequências genéticas não apresentam similaridade alguma com as já depositadas nas bases de dados mundiais e são comumente denominadas de “no hits”. Conceitos recentes, denominam esses “no hits” como “genes órfãos” e postulam que o surgimento dos mesmos são resultantes de respostas adaptativas específicas de cada espécie em função de estresses e condições adversas por essas enfrentadas, durante o processo evolutivo. Nosso trabalho está focado na identificação e caracterização funcional dos genes órfãos de café, por apresentarem alto potencial de inovação e aplicação biotecnológica. O presente estudo apresenta dados obtidos para alguns desses genes órfãos, denominados de *CcUnk* (Unknown), previamente identificados no genoma de *C. canephora* e com foco especial àqueles *CcUnk* genes potencialmente envolvidos na resposta à seca.

PALAVRAS-CHAVE: *Coffea canephora*, genes órfãos, estresse abiótico, expressão gênica.

IDENTIFICATION AND CHARACTERIZATION OF OPHAN GENES (“NO HITS”) IN COFFEE (*COFFEA CANEPHORA*), INVOLVED IN DROUGHT-STRESS RESPONSES

ABSTRACT: Coffee is a perennial crop considered one of the most important agricultural commodities in the world. By consequence, aiming at the establishment of tools to help accelerate the genetic improvement of this species, significant advances in coffee genomics have occurred in recent years. As an example, one can cite the recent completion of the complete genome sequencing of *Coffea canephora*, which will serve as a reference work for use in advanced molecular genetics, applied directly to the genetic improvement of this species, such as the establishment of genome-wide selection programs (GWS) in coffee. Recent bioinformatic analyzes of complete genomes of plants indicate that about 20-30% of the total complete set of genes are novel and specific to each species. That is, these genomic sequences do not exhibit any similarity with those already deposited in global databases and are commonly called “no hits”. Recent concepts, called these “no hits” as “orphan genes” and postulate that the emergence of these are the result of adaptive responses specific to each species as a function of stresses and adverse conditions faced by these plants during the evolutionary process. Our work is focused on the identification and functional characterization of orphan genes from coffee, which show a high potential for innovation and biotechnological applications. This study presents data obtained for some of orphan genes, called *CcUnk* (Unknown), previously identified in the genome of *C. canephora* with special focus on *CcUnk* genes potentially involved in drought-responses.

KEY WORDS: *Coffea canephora*, orphan genes, abiotic stress, gene expression.

INTRODUÇÃO

O café, trata-se de uma cultura perene e de período juvenil longo, tornando o melhoramento genético através de técnicas convencionais lento (BORÉM; ALMEIDA, 2011). Sendo assim, as técnicas de biologia avançada tais como a genômica estrutural e funcional, a proteômica, a bioinformática e a prospecção e transformação genética, podem oferecer alternativas para se reduzir o tempo de obtenção de novos genótipos e atuar como uma ferramenta auxiliar nos

programas de melhoramento genético (LASHERMES *et al.*, 2002). Um requisito primordial para a aplicação destas técnicas biotecnológicas é a identificação e a caracterização funcional de genes com potencial utilização nestes processos (HORAN *et al.*, 2008). Desta forma, a realização de uma análise exploratória e discriminatória dos genes ou fatores genéticos, associados com uma determinada característica fenotípica de interesse agrônomo é de vital importância para a aplicação prática destas ferramentas no melhoramento genético do cafeeiro. A geração de bancos de dados de sequências nucleotídicas a partir do sequenciamento de genomas completos ou de projetos ESTs (“*Expressed Sequences Tags*”), dentre outros, proporcionam uma fonte de determinantes genéticos para a busca de alternativas no controle de pragas e doenças, bem como no aumento do rendimento e adaptabilidade das espécies agrônomicas. Resultados de análises de bioinformática (MONDEGO *et al.*, 2011; VIEIRA *et al.*, 2006) mostram que aproximadamente 30% dessas sequências não possuem nenhuma similaridade, (sequências denominadas *no hits*) com aquelas conhecidas no GenBank¹. Horan *et al.* (2008) identificaram e caracterizaram os genes que codificam proteínas de função desconhecida no genoma de *Arabidopsis thaliana*. Outros trabalhos também têm identificado *no hits*, como em Alvarenga *et al.* (2010) e Vinecky *et al.* (2012), que identificaram nas bases de dados do PGBC *no hits* potencialmente relacionados a resistência à doenças e à seca, em café. Estes *no hits* podem ser genes que desempenham papéis cruciais em determinados processos biológicos do cafeeiro, visto que a caracterização destes genes desconhecidos é o foco deste trabalho.

MATERIAL E MÉTODOS

Identificação dos genes órfãos: Como resultado do Consórcio Internacional de Sequenciamento do Genoma de *Coffea canephora*, foram preditos cerca de 25 mil genes (WINCKER *et al.*, 2011- <http://coffee-genome.org>; acesso restrito). Após o fluxograma de análises de identificação de genes ortólogos (*Arabidopsis*, videira e tomate), com o OrthoMCL2 (versão 1.4), cerca de 5 mil genes (~20%) de *C. canephora* não apresentaram nenhum ortólogo entre os genomas analisados, sendo potenciais genes órfãos de café. Como forma de priorizar os trabalhos, estas 5 mil sequências foram utilizadas para a busca por similaridade nas bases de dados do NCBI. As análises foram realizadas por meio da ferramenta BLASTp. As sequências foram contrastadas com o banco de dados do NCBI NR no GenBank utilizando o software Geneious (KEARSE *et al.*, 2012). Foram considerados “*no hits*” os resultados de BLASTp ($E\text{-value} \geq 1e^{-1}$), que não apresentaram similaridade com aquelas conhecidas nos bancos de dados do GenBank, sendo estas sequências, posteriormente utilizadas, para análises de expressão *in silico*.

Análises da expressão gênica (in silico e qPCR): Os dados utilizados neste estudo foram obtidos por meio de dois métodos de sequenciamento: a) dados gerados com a tecnologia de sequenciamento do tipo Sanger, provenientes das bibliotecas relatadas em Vieira *et al.* (2006), Poncet *et al.* (2006) e Lin *et al.* (2005); b) dados de sequências obtidas com a tecnologia de sequenciamento em larga escala do tipo 454 da plataforma Roche (GS-FLX Titanium strategy), dados não publicados. As sequências 454 Roche foram geradas a partir de extratos de RNA de raízes de *C. canephora* dos clones 14, 73 e 120 considerados tolerantes à seca, e clone 22 considerado sensível à seca, nas condições de irrigação (I) e não-irrigação (NI) (MARRACCINI *et al.*, 2012). As análises de expressão *in silico* foram realizadas separadamente, conforme as tecnologias de sequenciamento utilizadas. Nestas análises (*Northern* eletrônico) foi utilizado o pacote DNASTar, software Qseq, versão 7.0, para mapeamento dos *no hits* encontrados na coleção genômica de 25 mil genes de *C. canephora*, gerando assim os dados de expressão dos *no hits* em cada biblioteca. Os dados foram previamente normalizados com a padronização da expressão dos *no hits* sempre relativos à expressão do gene da ubiquitina (*CcUBQ10*), usado como um calibrador interno de referência. As amostras e métodos utilizados nas análises de qPCR foram as mesmas descritas por Vieira *et al.*, 2013.

Caracterização funcional: Para a caracterização funcional dos genes órfãos de café, foi utilizado o sistema de transformação genética em *Setaria viridis* via *Agrobacterium tumefaciens*. O protocolo utilizado para transformação de *S. viridis* foi conforme descrito por Brutnell *et al.* (2010). Para transformação foram usados calos embriogênicos obtidos a partir de sementes maduras de *S. viridis* (acesso A10). Os calos embriogênicos foram incubados com a suspensão de *A. tumefaciens* (OD₆₀₀=0,6) contendo a construção UBI::CcUNK8 por 5 min. e co-cultivados no escuro a 22 °C por três dias, em meio de indução de calos (CIM) acrescido de 200 µM de acetoseringona. Ao final do co-cultivo, os calos foram transferidos para o meio CIM suplementado com 150 mg/L de Timentin® onde permaneceram por uma semana no escuro com temperatura de 25 ± 2 °C. Após esse período, os mesmos foram transferidos para meio seletivo CIM contendo 150 mg/L de Timentin® e 30 mg/L do agente seletivo Higromicina B, por mais uma semana no escuro a 25 ± 2 °C. Os calos que sobreviveram a essa primeira seleção foram transferidos para o meio de regeneração (MRS), acrescido de 150 mg/L de Timentin® e 30 mg/L Higromicina B. Nesse momento os calos foram acondicionados sob fotoperíodo de 16 horas luz e temperatura de 25 ± 2 °C. As plântulas regeneradas com parte aérea e raízes bem desenvolvidas foram transferidas para Magentas™, contendo sais e vitaminas MS (MURASHIGE & SKOOG, 1964) suplementado com 150 mg/L de Timentin® e 30 mg/L Higromicina B. As plântulas com aproximadamente 10 cm de parte aérea e com raízes principais e adventícias bem desenvolvidas, foram transferidas para copos plásticos de 100 mL com a mistura de terra, substrato Plantmax® e vermiculita (3:3:1). Os mesmos foram cobertos com sacos plásticos e

¹ Disponível em: ><http://blast.ncbi.nlm.nih.gov/Blast.cgi><

aclimatados em câmara Fitotron com temperatura 25 ± 2 °C, umidade relativa 65% e intensidade luminosa de 400 $\mu\text{M}/\text{m}^2/\text{s}$.

RESULTADOS E DISCUSSÃO

Na figura 1 estão apresentados alguns resultados de *Northern* eletrônico (Heat-Map / TreeView), indicando alguns genes órfãos com expressão diferencial nos tecidos analisados (calos, frutos e folhas-Figura 1A), assim como em folhas controle e sob estresse hídrico (Figura 1B).

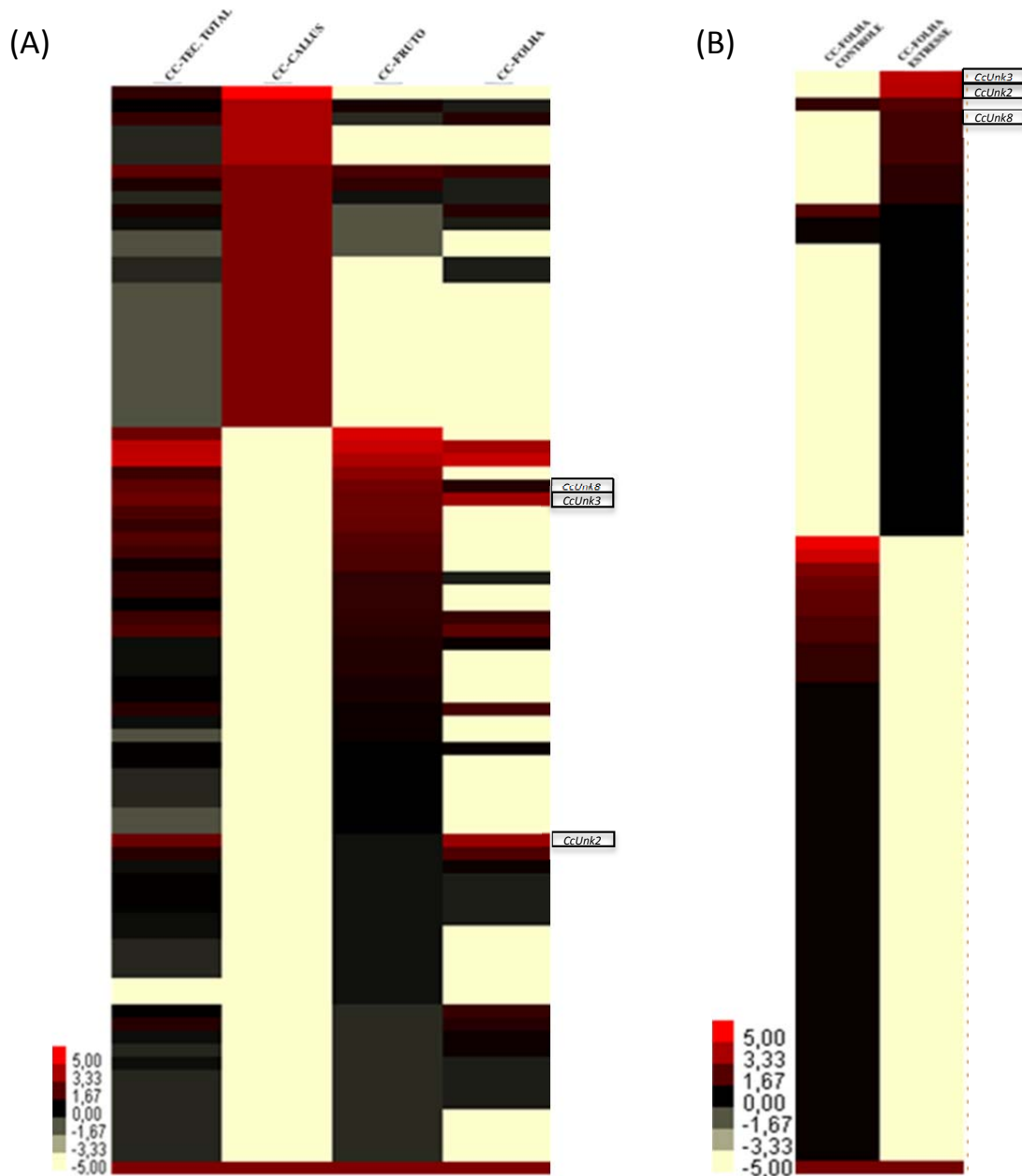


Figura 1: Expressão gênica *in silico* dos *no hits* nas bibliotecas de (A) diferentes tecidos de *C. canephora* e (B) bibliotecas de folhas controle e sob estresse de seca, produzidas utilizando a tecnologia Sanger.

Quando comparamos as bibliotecas de folhas controle e de plantas submetidas ao estresse de seca (Figura 1B), pôde-se observar a expressão diferencial de vários *CcUnk* genes. É possível ressaltar, como exemplo, o caso do *CcUnk2*, *CcUnk3* e *CcUnk8*, que parecem ser altamente induzidos com o estresse de seca. A expressão destes três *CcUnk* genes avaliada por qPCR está apresentada na figura 2. Para esses estudos foram utilizadas folhas de clones de *C. canephora* tolerantes à seca 14, 73 e 120 e o clone 22 sensível à seca. Os clones analisados foram submetidos ao estresse de seca (não irrigado, NI) e em condições controle (irrigado, I). Observa-se pelos dados apresentados na figura 2, que todos os genes *CcUnk* avaliados apresentaram perfis de expressão diferencial para as condições hídricas avaliadas, assim como comportamento diferencial em função do clone analisado, confirmando os resultados obtidos pelas análises *in silico*.

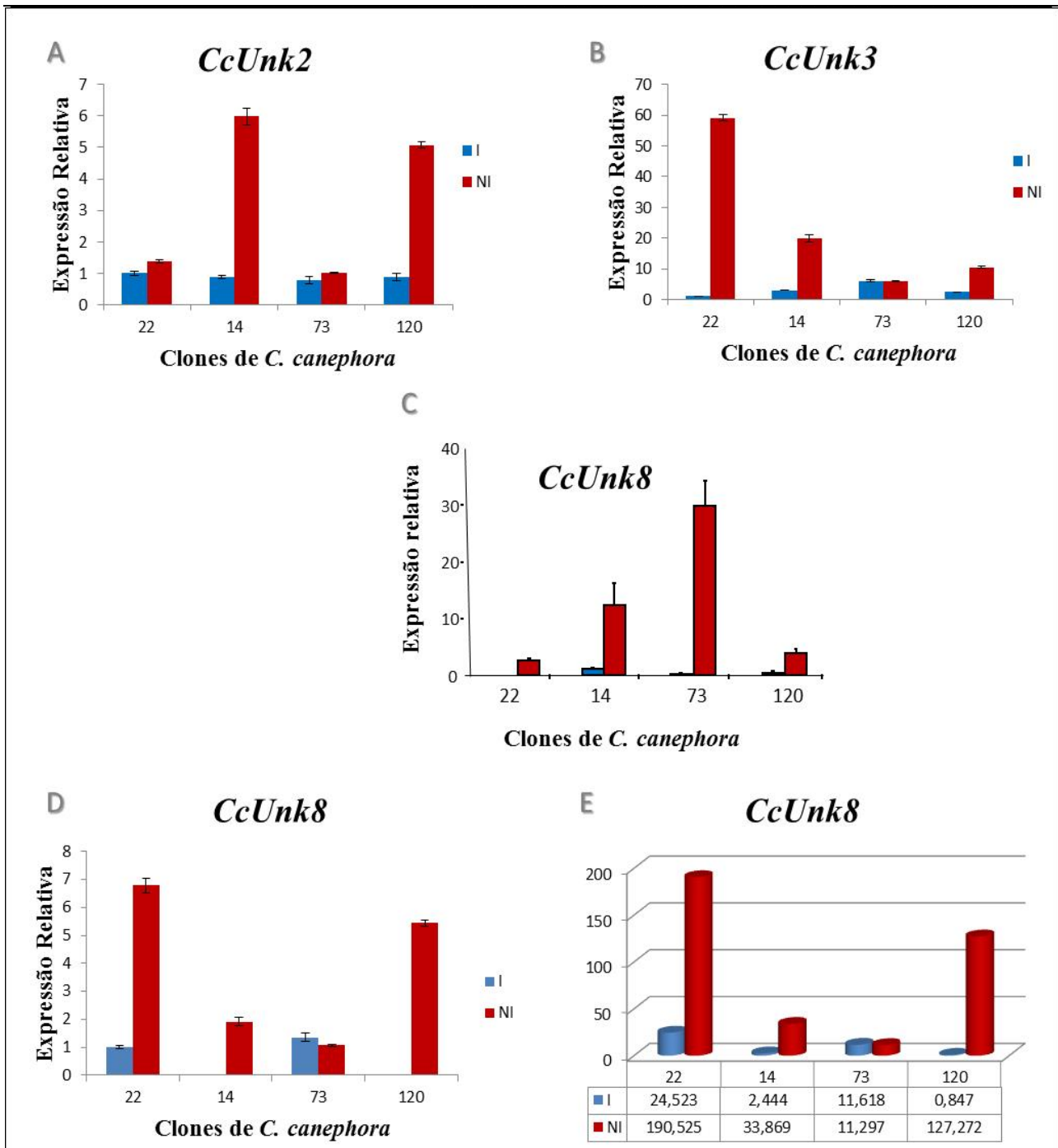


Figura 2: Expressão dos genes candidatos *CcUnk3*, *CcUnk2* e *CcUnk8* analisada por qPCR em tempo real, a partir de RNA extraído de folhas (A, B, C) e raízes (D) dos clones 22, 14, 73 e 120 de *C. canephora* submetidos ao estresse de seca (NI: Não irrigado) e sem estresse (I: Irrigado). O gene *CcUBQ10* foi utilizado como controle endógeno. O calibrador interno utilizado foi o 22I. A figura 2E, apresenta o resultado da análise *in silico* da expressão do gene *CcUnk8* a partir de dados de sequenciamento em larga escala (454 Roche) normalizados por RPKM.

A figura 3 apresenta uma representação esquemática da construção *Ubi::CcUnk8*, assim como um esquema dos passos utilizados na estratégia de transformação em *S. viridis*, visando-se a caracterização da função do gene *CcUnk8* e a comprovação de seu envolvimento na tolerância à seca em plantas. Plantas transgênicas de *S. viridis* já transformadas com a construção *Ubi::CcUnk8* já foram obtidas (Figura 4) e estão em processo de avaliação.

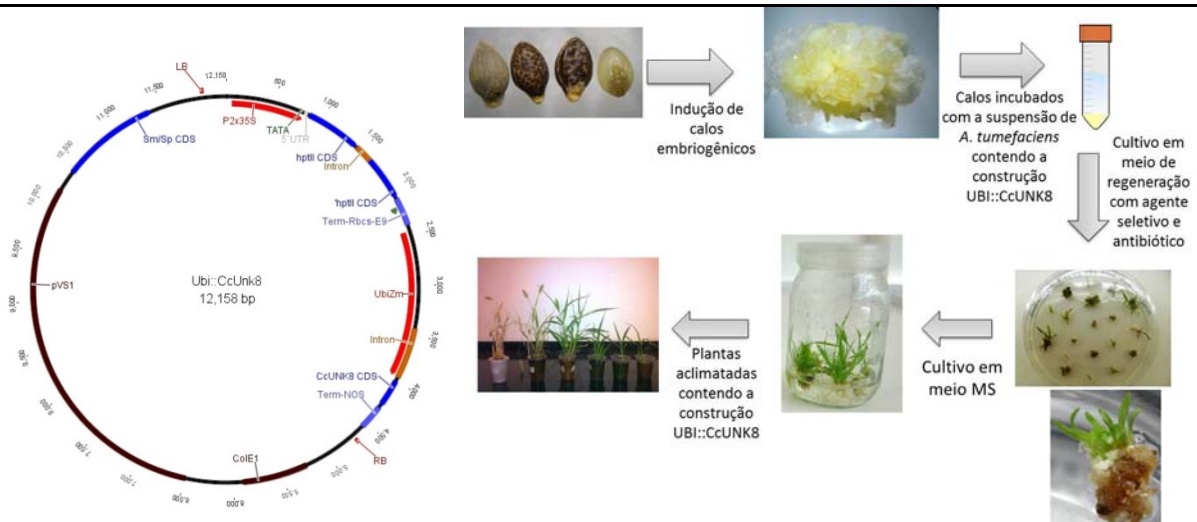


Figura 3: Expressão dos genes candidatos *CcUnk3* (A), *CcUnk2* (B) e *CcUnk8* (C) analisada por qPCR em tempo real, a partir de RNA extraído de folhas dos clones 22, 14, 73 e 120 de *C. canephora* submetidos ao estresse de seca (NI: Não irrigado) e sem estresse (I: Irrigado). O calibrador interno utilizado foi o 22I. O gene *CcUBQ10* foi utilizado como controle endógeno.

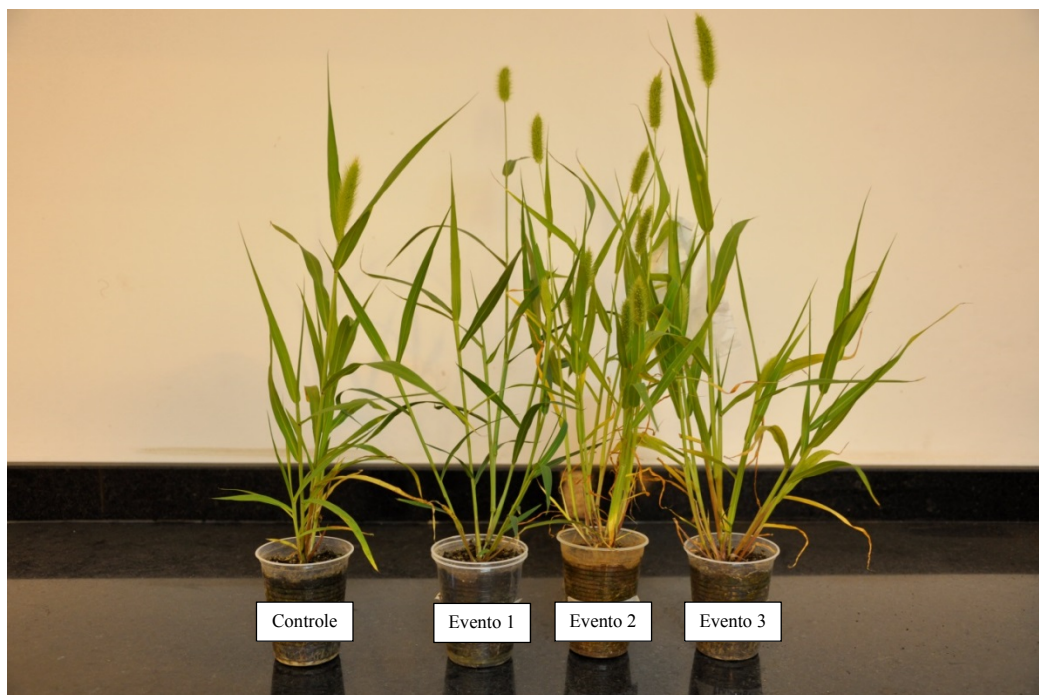


Figura 4: Plantas transgênicas de *S. viridis*, transformadas com a construção *Ubi::CcUnk8*, em fase de avaliação.

CONCLUSÕES

Os genes órfãos (*no hits*) de café foram identificados e caracterizados *in silico* de acordo com o tecido em que foram expressos na espécie *C. canephora*. Foi possível identificar e caracterizar *CcUnk* genes que parecem responder de forma significativa quanto a tolerância e suscetibilidade ao estresse de seca. Os resultados apresentados neste trabalho proporcionam o primeiro estudo sistemático de caracterização funcional de genes órfãos de café, provendo insumos com alto potencial inovador para trabalhos biotecnológicos em plantas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVARENGA, S. M., *et al.* *In silico* identification of coffee genome expressed sequences potentially associated with resistance to diseases. **Genetics and Molecular Biology**, v. 33, n. 4, p. 795-806, Jul 2010.

- BORÉM, A.; ALMEIDA, G. D. **Plantas Geneticamente modificadas: Desafios e oportunidades para Regiões Tropicais**, 1ª ed, Editora UFV, Viçosa, 2011.
- BRUTNELL, T.P.; WANG, L.; SWARWOOD, K.; GOLDSCHMIDT, A.; JACKSON, D.; ZHU, X.G.; KELLOGG, E.; ECK, J.V. *Setaria viridis*: a model for C4 photosynthesis. *The plant cell*. 2010.
- EISEN, M. B., *et al.* Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 25, p. 14863-14868, 1998.
- HORAN, K., *et al.* Annotating genes of known and unknown function by large-scale coexpression analysis. **Plant Physiology**, v. 147, n. 1, p.47-51, 2008.
- KEARSE, M., *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647-1649, 2012.
- LASHERMES P, ANDRADE AC, ETIENNE H. 2008. Genomics of coffee, one of the world's largest traded commodities. In: Moore H, Ming R, eds. **Genomics of tropical crop plants**. Berlin: Springer, 203–226.
- LIN, C. W., *et al.* Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. **Theoretical and Applied Genetics**, v. 112, n. 1, p. 114-130, 2005.
- MARRACCINI, P., *et al.* Differentially expressed genes and proteins upon drought stress in tolerant and sensitive genotypes of *Coffea canephora*. **Journal of Experimental Botany**, v. 63, n. 11, p. 4191-4212, 2012.
- MONDEGO, J. M. C., *et al.* An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. **BMC Plant Biology**, v. 11: 30, 2011.
- PONCET, V., *et al.* SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular Genetics and Genomics**, v. 276, n. 5, p. 436-449, 2006.
- VIEIRA, N. G. ; ALVES, G. S. C. ; FREIRE, L. P. ; CARNEIRO, F. A. ; MENDONÇA, L. M. C. ; COSTA, T. S. ; ALEKCEVETCH, J. C. ; MELO, J. A. T. ; BLOCH JR, C. ; MARRACCINI, P. ; ANDRADE, A. C. . NO-HITOMICS: UNDERSTANDING THE COFFEE UNKNOWNNS. In: **The 24th International Conference on Coffee Science**, San José, Costa Rica: ICAFE, . p. 223, 2012.
- VIEIRA N.G., *et al.* Different molecular mechanisms account for drought tolerance in *Coffea canephora* var. Conilon. **Tropical Plant Biology**, online, DOI: 10.1007/s12042-013-9126-0, 2013.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 12, p. 5463-5467, 1977.
- VIEIRA, L. G. E., *et al.* Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 95-108, Mar 2006.
- VINECKY, F. ; DA SILVA, F. R. ; ANDRADE, A. C. Análise *in silico* das bibliotecas de cDNA SH2 e SH3 para a identificação de genes responsivos à seca em cafeeiro. **Coffee Science**, v. 7, p. 1-19, 2012.
- WINCKER, P., *et al.* Sequencing the coffee genome. In: **XIX Plant and Animal Genome Conference**, San Diego, California, n. 152, p. 15-19, 2011.