



NATALIA GOMES VIEIRA

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE
GENES ÓRFÃOS (“NO HITS”) DE CAFÉ
(*COFFEA* spp.)**

LAVRAS - MG

2013

NATALIA GOMES VIEIRA

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES ÓRFÃOS (“NO
HITS”) DE CAFÉ (*COFFEA* spp.)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, para obtenção do título de Mestre.

Orientador

PhD Alan Carvalho Andrade

Coorientador

PhD Pierre Marraccini

LAVRAS - MG

2013

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Vieira, Natalia Gomes.

Identificação e caracterização de genes órfãos (“*no hits*”) de
café (*Coffea* spp.) / Natalia Gomes Vieira. – Lavras : UFLA, 2013.
130 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2013.
Orientador: Alan Carvalho Andrade.
Bibliografia.

1. Cafeeiro. 2. Estresse. 3. Expressão gênica. 4. Base de dados. I.
Universidade Federal de Lavras. II. Título.

CDD – 631.523

NATALIA GOMES VIEIRA

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES ÓRFÃOS (“NO
HITS”) DE CAFÉ (*COFFEA* spp.)**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, para obtenção do título de Mestre.

APROVADA em 28 de fevereiro de 2013.

Dr. Alan Carvalho Andrade

EMBRAPA Cenargen

Dr. Hugo Bruno Correa Molinari

EMBRAPA Agroenergia

Dr. Pierre Roger Marraccini

CIRAD

PhD Alan Carvalho Andrade

Orientador

LAVRAS - MG

2013

Aos amantes da ciência, que assim como eu se fascinam pelo exuberante mundo da pesquisa molecular.

Aos meus pais; Iara e Antônio Carlos, pela importância incondicional que têm na minha vida, sempre abdicando de suas próprias vontades para que meus sonhos e objetivos se realizem.

DEDICO

AGRADECIMENTOS

Primeiramente a Deus, que ilumina meus passos, me ama pelo que sou e me fortalece diante das adversidades impulsionando-me a conquistar lugares altos;

A minha família, em especial meus pais; Antônio Carlos e Iara, e meus irmãos Diogo e Ana Gabriella, pela confiança, apoio, dedicação e amor. Sem vocês não seria possível concretizar esse trabalho!

As minhas amigas de infância e de sempre; Jú e Cau, pela amizade verdadeira, pelo apoio em minhas decisões, conselhos e por serem tão especiais;

Aos meus amigos e companheiros eternos, que conheci no laboratório, mas que pelo curso da vida não estão mais lá, pelos momentos de descontração e amizade, Kelly, Éder, Mônica, Gabriel, Mari, Ingrid, Fabrício, Bárbara, Maria, Marry, André, e em especial agradeço ao Felipe e a Lú, pela bela amizade construída, ensinamentos, conselhos e dedicação. Vocês são demais!

Aos amigos do mestrado ou que conheci em Lavras que são agora amigos para vida toda, Marcelo, Lara e Mayara, pela torcida e por dividir comigo as mesmas angústias e principalmente pela amizade e muitos momentos agradáveis e descontraídos.

As minhas amigas de república, Ananza, Noelly e Laiane, obrigada pela oportunidade de conhecer pessoas tão diferentes de mim e especiais. Em especial agradeço à minha amiga Michelle, pelo apoio independente de tudo e de todos, pelos momentos de estudo, descontração, espontaneidade, confiança, amor, confidência, conselhos, enfim, você é ótima!

A todos os meus amigos da Embrapa, dentre eles, Fernanda, Tati, Jorge, Rafa, Jean, Érica e Edriana, por todo trabalho feito para que eu pudesse realizar o meu, e que nunca mediram esforços para poder me ajudar. Muito obrigada!

Ao meu namorado Kollien, pelo amor, carinho, paciência, cumplicidade e compreensão. Sem o seu apoio a minha vida não seria a mesma!

Ao Dr. Alan, pelos ensinamentos e orientação, pois não são somente dois anos de mestrado, mas sete anos de convivência, confiança, apoio e amizade. Ofereceu-me a oportunidade de conhecer a ciência pela qual me apaixonei, e tem contribuindo de forma significativa para minha capacitação profissional;

Ao Dr. Pierre, pois com ele aprendi o verdadeiro significado da palavra humildade. Sempre disponível para me ajudar e me ensinar com paciência e dedicação. Um profissional igual a você impulsiona a qualquer pessoa a seguir adiante!

Aos professores do curso de Biotecnologia Vegetal, pelos valiosos ensinamentos, em especial ao professor Luciano, Coordenador do curso de Biotecnologia Vegetal, pela oportunidade;

Aos membros da banca examinadora, Dr. Alan Andrade, Dr. Pierre Marraccini e Dr. Hugo Molinari, por aceitarem o convite para fazer parte da banca e contribuir construtivamente com o meu trabalho;

Enfim, a todos que de alguma forma, direta ou indiretamente contribuíram para a realização desse trabalho. Obrigada!

“A atividade de pensar confere ao homem asas para mover-se no mundo e raízes para aprofundar-se na realidade.”

René Descartes

RESUMO

O café é uma das *commodities* agrícolas mais importantes mundialmente. Programas de melhoramento genético que visam a introdução de caracteres para obtenção de novas variedades com características agronômicas superiores, como floração uniforme, rendimento, tamanho do grão, qualidade da bebida, menor teor de cafeína e resistência à doenças e à seca estão em andamento à décadas. Contudo, com um número considerável de sequências gênicas disponíveis nos bancos de dados *Coffea* spp., a identificação e a caracterização de novos genes com função desconhecida (*no hits*) é viável e indispensável. O objetivo com este trabalho consistiu em identificar e caracterizar funcionalmente, por meio de análises *in silico* e por qPCR da expressão gênica, *no hits* encontrados nas bases de dados do café. Para as análises *in silico* os *no hits* foram identificados por meio da busca por similaridade nos bancos de dados do NCBI. Bibliotecas de *Coffea arabica* e *Coffea canephora* de diferentes tecidos foram geradas a partir dos dados de sequenciamento utilizando o método Sanger e o método 454. Foi realizado um *Northern* eletrônico dos *no hits* nas diferentes bibliotecas de café. Para os ensaios de qPCR o material vegetal utilizado foi proveniente de diferentes tecidos (folhas, raízes, frutos e meristemas) das espécies de *C. arabica* e *C. canephora*, submetidos à diferentes condições biológicas (estresse biótico e abiótico) e frutos em estádios de desenvolvimento inicial, intermediário e tardio. Mil quinhentos e sessenta e oito genes órfãos foram identificados e caracterizados nas análises *in silico*, de acordo com as bibliotecas em que foram expressos. Com isso, foram selecionados e testados em qPCR, dezessete genes órfãos potencialmente envolvidos na tolerância a estresses bióticos e abióticos, assim como genes órfãos implicados no desenvolvimento do fruto de café. Estes genes responderam diferencialmente nos tecidos e condições biológicas analisadas, corroborando com as análises *in silico* previamente apresentadas e oferecendo indícios de uma provável participação em processos biológicos específicos do cafeeiro.

Palavras-chave: Cafeeiro. Estresse. Base de dados. Expressão gênica.

ABSTRACT

Coffee is one of the most important agricultural commodities worldwide. Breeding programs aimed at introducing new characters to produce novel varieties with superior agronomic traits, such as, uniform flowering, yield, grain size, beverage quality, lower caffeine content, as well as, disease and drought resistance have been ongoing for several decades. However, with a large number of gene sequences available in *Coffea* spp. databases, the identification and characterization of novel genes with unknown function (*no hits*) are feasible and necessary. The aim of this study was to identify and functionally characterize *no hits* found in coffee databases, using *in silico* analysis and qPCR gene expression. For *in silico* analyses, *no hits* were identified by similarity search in NCBI databases. *Coffea arabica* and *Coffea canephora* libraries from different tissues were generated by sequencing data using the Sanger and 454 methods. We performed an electronic Northern search of the *no hits* in the coffee libraries. Plant material, for qPCR assays, was obtained from different tissues (leaves, roots, fruits and meristems) of the species *C. arabica* and *C. canephora* subjected to different biological conditions (biotic and abiotic stresses) and fruits at early, intermediate and late development stages. As a result, *in silico* analyzes identified and characterized 1.568 orphan genes in accordance with the libraries in which they were expressed. Therefore, we selected and performed qPCR of 17 orphan genes potentially involved in tolerance to biotic and abiotic stresses, as well as, orphan genes involved in coffee fruit development. These genes responded differentially in tissues and with biological conditions, confirming *in silico* analyzes previously obtained and providing evidences for their probable involvement in specific biological processes in coffee plants.

Keywords: Coffee. Stress. Databases. Gene expression.

LISTA DE FIGURAS

CAPÍTULO 1

- Figura 1 Origem geográfica dos principais grupos genéticos de *C. canephora*. Vermelho: origem geográfica do grupo Guineano, verde: origem geográfica do grupo genético congolês; círculos com o nome dos grupos genéticos: Origem geográfica de cada grupo; áreas hachuradas: refúgios do último máximo glacial (20.000 anos atrás). Fonte: Montagnon, Cubry e Leroy (2012)..... 30
- Figura 2 Gráfico das alterações teciduais que ocorrem durante o desenvolvimento dos frutos de café e maturação. e: endosperma; ps: perisperma; pc: pericarpo. Fonte: Salmona et al. (2008), adaptado de Castro e Marraccini (2006)..... 37
- Figura 3 Passos gerais envolvidos na análise de ESTs. 1. Sequências primárias de ESTs são verificadas quanto à presença do vetor, regiões repetitivas, que são excisadas ou mascaradas. Sequências com baixa qualidade, singletons e sequências muito curtas também são removidas. 2. ESTs são então agrupadas e montadas para gerar sequências consenso (“transcritos putativos”). 3. Buscas por similaridade em banco de dados de DNA são realizadas para identificar homólogos e atribuir uma possível função 4. Putativos peptídeos são obtidos por tradução das sequências consenso. 5. Buscas por similaridade no banco de dados de proteínas são realizadas para atribuir a(s) função (ões) putativa(s). A análise é estendida pela anotação funcional seguida de visualização e interpretação dos resultados. Adaptado de Nagaraj, Gasser e Ranganathan (2007)..... 38
- Figura 4 Perfis de expressão dos genes *CcUNK8*, *CcUNK1* e *CcUNK10* nas folhas dos clones 14, 22 e 120 de *C. canephora* submetidos (NI: Não irrigado) ou não (I: Irrigado) à seca. Resultados da expressão usando o 14I como calibrador interno. A expressão do gene *CcUBQ10* foi utilizada como referência para a quantificação relativa, exceto para o gene marcado com asterisco que foi usado o gene *CcGAPDH* como referência. Adaptado de Marraccini et al. (2012)..... 46
- ### CAPÍTULO 2
- Figura 1 Representação esquemática de cada locus dos *nos hits* encontrados, distribuídos nos 11 cromossomos de café..... 69

Figura 2	Expressão gênica <i>in silico</i> dos <i>no hits</i> nas bibliotecas de <i>C. arabica</i> (CA) utilizando a tecnologia Sanger. (*) <i>no hits</i> citados no texto. (*) <i>no hits</i> citados no texto e selecionados para qPCR.....	72
Figura 3	Expressão gênica <i>in silico</i> dos <i>no hits</i> nas bibliotecas de <i>C. canephora</i> (CC) utilizando a tecnologia Sanger. (*) <i>no hits</i> citados no texto. (*) <i>no hits</i> citados no texto e selecionados para qPCR.....	74
Figura 4	Expressão gênica <i>in silico</i> dos <i>no hits</i> nas bibliotecas de folhas de <i>C. canephora</i> (CC) com e sem estresse de seca, utilizando a tecnologia Sanger. À direita, um enfoque nos genes mais expressos. (*) genes citados no texto. (*) genes citados no texto e selecionados para qPCR.....	76
Figura 5	Expressão gênica dos <i>no hits</i> nas bibliotecas de frutos de <i>C. canephora</i> (CC) em diferentes estádios de desenvolvimento, utilizando a tecnologia Sanger. (*) <i>no hits</i> citados no texto. (*) <i>no hits</i> citados no texto e selecionados para qPCR.....	78
Figura 6	Expressão gênica dos <i>no hits</i> nas bibliotecas de raiz de <i>C. canephora</i> (CC) utilizando a tecnologia 454. (*) <i>no hits</i> citados no texto. (*) <i>no hits</i> citados no texto e selecionados para qPCR.....	81
Figura 7	Expressão gênica dos <i>no hits</i> nas bibliotecas de meristemas de <i>C. arabica</i> (CA) utilizando a tecnologia 454. (*) <i>no hits</i> citados no texto. (*) <i>no hits</i> citados no texto e selecionados para qPCR.....	84
CAPÍTULO 3		
Figura 1	Curva de dissociação da qPCR para os <i>primers</i> 42747 (A) e 36640 (B) utilizando o mix dos estádios de desenvolvimento para os diferentes tecidos, perisperma, endosperma e pericarpo de frutos de <i>C. canephora</i>	104
Figura 2	Perfis de expressão dos <i>no hits</i> 33590 (A), 30926 (B) e 33656 (C) nas raízes dos clones 14 e 22 de <i>C. canephora</i> submetidos (raiz com nematoide) ou não (raiz controle) ao estresse biótico. O calibrador interno utilizado foi o 14 controle (sem infecção de nematoide). O gene <i>CcUBQ10</i> foi utilizado como controle endógeno.....	106

- Figura 3 Perfis de expressão dos genes candidatos 33656 (A) e 30926 (B) nas 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..... 109
- Figura 4 Expressão do *no hit 33656* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 33656* no mix dos estádios de desenvolvimento nos diferentes tecidos perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 33656* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 33656* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 33656* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 33656* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador interno, em (B), (C), (D) e (E) o calibrador interno utilizado foi o 90PE..... 112
- Figura 5 Expressão do *no hit 07128* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 07128* no mix dos estádios de desenvolvimento nos diferentes tecidos perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador, em (B), (C), (D) e (E) o calibrador utilizado foi o 210PE..... 114

Figura 6 Expressão do *no hit 18112* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 18112* no mix dos estádios de desenvolvimento nos diferentes tecidos perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador interno, em (B), (C), (D) e (E) o calibrador interno utilizado foi o 210PE..... 116

Figura 7 Expressão do *no hit 02344* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 02344* no mix dos estádios de desenvolvimento nos diferentes tecidos perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador interno, em (B), (C), (D) e (E) o calibrador interno utilizado foi o 90PE..... 118

- Figura 8 Expressão do *no hit 24645* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 24645* no mix dos estádios de desenvolvimento nos diferentes tecidos perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 24645* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 24645* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 24645* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 24645* nos estádios de desenvolvimento no pericarpo. O gene *CcUBI10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador, em (B), (C), (D) e (E) o calibrador utilizado foi o 90PE..... 120
- Figura 9 Perfis de expressão dos *no hits 14149* (A, B), *05517* (C, D), *34630* (E, F), *14549* (G, H) e *42816* (I, J) na qPCR e *in silico*, respectivamente, nas raízes dos clones de *C. canephora* 14, 73 e 120 (tolerantes à seca) e o clone 22 (sensível à seca) submetidos ao estresse hídrico (NI: não irrigado) e sem estresse (I: Irrigado). Para as reações de qPCR o controle endógeno utilizado foi o *CcUBQ10*. O calibrador utilizado para normalização dos dados da qPCR foi o 22I. Nas análises *in silico* foi utilizado o BLASTp contra o banco de dados NR do NCBI, filtrados com E-value $\leq 1e-1$ 123
- Figura 10 Perfis de expressão dos genes candidatos *11524* (A), *17768* (B), *25639* (C), *32791* (D) e *30926* (E) nos meristemas das cultivares Rubi e I59 de *C. arabica* submetidos ao estresse de seca (I: Irrigado) e sem estresse (NI: não irrigado). Rubi NI como calibrador em A e B e Rubi I como calibrador em C, D e E. O gene *CcUBQ10* foi utilizado como controle endógeno.. 127

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1	Descrição das bibliotecas de ESTs geradas pelo Projeto Brasileiro do Genoma Café. Fonte: Vieira et al. (2006).....	41
----------	--	----

CAPÍTULO 2

Tabela 1	Bibliotecas utilizadas nas análises de expressão <i>in silico</i> . Os dados gerados utilizando a tecnologia Sanger.....	64
Tabela 2	Subdivisões das bibliotecas CC-Folhas e CC- Fruto	66
Tabela 3	Bibliotecas utilizadas nas análises de expressão <i>in silico</i> . Os dados gerados utilizando a tecnologia 454.....	67

CAPÍTULO 3

Tabela 1	Material vegetal utilizado para a extração de RNA e condução dos experimentos de qPCR.....	91
Tabela 2	Separação dos tecidos embrionários dos frutos de <i>C. canephora</i> (clone, linha 6, planta 35) coletados em diferentes estádios de desenvolvimento.....	95
Tabela 3	<i>Primers forward</i> (F) e <i>reverse</i> (R) utilizados para as análises de qPCR. NUBI corresponde ao gene endógeno <i>CcUBQ10</i> . Os <i>primers</i> estão identificados de acordo com a nomenclatura dos <i>no hits</i> encontrados nas análises <i>in silico</i> , sem o prefixo GSCOCT000 e o sufixo 001.....	99
Tabela 4	<i>Primers</i> dos <i>no hits</i> utilizados nos diferentes materiais vegetais analisados nos ensaios de qPCR.....	103

LISTA DE ABREVIATURAS

C.	<i>Coffea</i>
M.	<i>Meloidogyne</i>
est.	Estádio
Susp. Cel.	Suspensão celular
Tec.	Tecidos

LISTA DE SIGLAS

ABIC	Associação Brasileira da Indústria de Café
CENICAFE	Centro Nacional de Investigaciones de Café
BLAST	Basic Local Alignment Search Tool
CONAB	Companhia Nacional de Abastecimento
IAPAR	Instituto Agrônomo do Paraná
Incaper	Instituto Capixaba de Pesquisa e Assistência Rural
ICO	International Coffee Organization
IPCC	Intergovernmental Panel on Climate Change
IRD	Institut de Recherche pour le Développement
LGE	Laboratório de Genômica e Expressão
MAPA	Ministério da Agricultura Pecuária e Abastecimento
NCBI	National Center for Biotechnology Information
ONSA	Organização para Análise e Sequenciamento de Nucleotídeos
PBGC	Projeto Brasileiro do Genoma Café
SUCEST	Sugarcane Expressed Sequence Tag

LISTA DE SÍMBOLOS

Ψ	potencial hídrico
$^{\circ}\text{C}$	graus Celsius
%	percentage
®	Registrado
TM	Trade Market
x	Vezes
C_T	Ciclo <i>Threshold</i>
pH	potencial hidrogeniônico
MPa	megapascal
mL	mililitro
mM	milimolar
nm	nanômetro
ng	nanograma
μg	micrograma

SUMÁRIO

	CAPÍTULO 1 INTRODUÇÃO GERAL.....	20
1	INTRODUÇÃO.....	21
2	REFERENCIAL TEÓRICO.....	24
2.1	O café.....	24
2.1.1	Caracterização, importância e economia.....	24
2.1.2	Melhoramento genético do cafeeiro.....	27
2.1.3	Estresse biótico e abiótico.....	32
2.1.4	Desenvolvimento do fruto.....	34
2.2	Bioinformática.....	37
2.2.1	Estudo de genomas expressos (EST) em plantas.....	37
2.2.2	Bases de dados do café.....	41
2.3	No hits.....	43
3	CONSIDERAÇÕES GERAIS.....	47
	REFERÊNCIAS.....	48
	CAPÍTULO 2 IDENTIFICAÇÃO E CARACTERIZAÇÃO DA EXPRESSÃO <i>IN SILICO</i> DOS <i>NO HITS</i> PRESENTES NAS BASES DE DADOS DE CAFÉ.....	58
1	INTRODUÇÃO.....	59
2	MATERIAL E MÉTODOS.....	61
2.1	Identificação dos <i>no hits</i>.....	61
2.1.1	Busca nas bases de dados e alinhamentos.....	61
2.2	Composição das bases de dados para as análises <i>in silico</i>.....	61
2.3	Bibliotecas geradas com o sequenciamento.....	62
2.3.1	Bibliotecas geradas com o sequenciamento do tipo sanger.....	62
2.3.2	Bibliotecas geradas com o sequenciamento do tipo 454.....	66
2.4	<i>Northern</i> eletrônico.....	68
3	RESULTADOS E DISCUSSÃO.....	69
3.1	Análise de expressão gênica <i>in silico</i> nas bibliotecas de <i>C. arabica</i> (CA) utilizando a tecnologia sanger.....	70
3.2	Análise de expressão gênica <i>in silico</i> das bibliotecas de <i>C. canephora</i> (CC) utilizando a tecnologia sanger.....	73
3.3	Análise de expressão gênica <i>in silico</i> das bibliotecas de raiz de <i>C. canephora</i> (CC) utilizando a tecnologia 454.....	79
3.4	Análise de expressão gênica <i>in silico</i> nas bibliotecas de meristema de <i>C. arabica</i> (CA) utilizando a tecnologia 454.....	82
4	CONCLUSÃO.....	85
	REFERÊNCIAS.....	86

CAPÍTULO 3 VALIDAÇÃO DE GENES CANDIDATOS POR MEIO DA ANÁLISE DA EXPRESSÃO GÊNICA POR PCR QUANTITATIVO EM TEMPO REAL (qPCR).....		88
1	INTRODUÇÃO.....	89
2	MATERIAL E MÉTODOS.....	91
2.1	Material vegetal.....	91
2.1.1	Meristemas de <i>C. arabica</i> cultivados no campo.....	92
2.1.2	Folhas e raízes de <i>C. canephora</i> dos clones 14, 22, 73 e 120 cultivados em casa de vegetação.....	92
2.1.3	Raízes de <i>C. canephora</i> inoculados com o nematoide <i>Meloidogyne incognita</i>.....	93
2.1.4	Frutos de <i>C. canephora</i> cultivados no campo.....	94
2.2	Extração de RNA e quantificação das alíquotas.....	96
2.3	Tratamento com DNase.....	96
2.4	Transcriptase reversa.....	97
2.5	Desenho de <i>primers</i> para as análises de expressão pela qPCR....	98
2.6	Avaliação da eficiência dos <i>primers</i>.....	100
2.7	Análises da expressão por PCR quantitativo em tempo real.....	100
3	RESULTADOS E DISCUSSÃO.....	102
3.1	Expressão dos genes <i>no hits</i> potencialmente envolvidos na resposta da planta ao estresse biótico.....	104
3.2	Efeito do estresse biótico sobre a expressão dos genes <i>no hits</i> em café.....	106
3.3	Expressão dos genes <i>no hits</i> potencialmente envolvidos nos mecanismos de resposta à seca nas folhas de café.....	107
3.4	Expressão do gene <i>33656</i> em folhas de <i>C. canephora</i> submetidos ao estresse de seca.....	107
3.5	Expressão do gene <i>30926</i> em folhas de <i>C. canephora</i> submetidos ao estresse hídrico.....	108
3.6	Expressão dos genes <i>no hits</i> potencialmente envolvidos no desenvolvimento do fruto de café.....	109
3.7	Expressão do gene <i>33656</i> em frutos de <i>C. canephora</i> em diferentes estádios de desenvolvimento e tecidos.....	111
3.8	Expressão do gene <i>07128</i> em frutos de <i>C. canephora</i> em diferentes estádios de desenvolvimento e tecidos.....	113
3.9	Expressão do gene <i>18112</i> em frutos de <i>C. canephora</i> em diferentes estádios de desenvolvimento e tecidos.....	115
3.10	Expressão do gene <i>02344</i> em frutos de <i>C. canephora</i> em diferentes estádios de desenvolvimento e tecidos.....	117

3.11	Expressão do gene <i>24645</i> em frutos de <i>C. canephora</i> em diferentes estádios de desenvolvimento e tecidos.....	119
3.12	Expressão dos genes <i>no hits</i> potencialmente envolvidos nos mecanismos de resposta à seca nas raízes de café.....	121
3.13	Expressão dos genes <i>no hits</i> potencialmente envolvidos nos mecanismos de resposta à seca nas raízes de café.....	121
3.14	Expressão do <i>no hit 14149</i> em resposta à seca nas raízes de <i>C. canephora</i>	122
3.15	Expressão do <i>no hit 34630</i> em resposta à seca nas raízes de <i>C. canephora</i>	124
3.16	Expressão do <i>no hit 14549</i> em resposta à seca nas raízes de <i>C. canephora</i>	124
3.17	Expressão do <i>no hit 42816</i> em resposta à seca nas raízes de <i>C. canephora</i>	124
3.18	Expressão dos genes <i>no hits</i> potencialmente envolvidos nos mecanismos de resposta à seca no meristema de café.....	125
3.19	Expressão dos <i>no hits 11524</i> e <i>17768</i> em resposta à seca nos meristemas de <i>C. canephora</i>	126
3.20	Expressão dos <i>no hits 25639</i> e <i>30926</i> em resposta à seca nos meristemas de <i>C. canephora</i>	126
3.21	Expressão do <i>no hit 32791</i> em resposta à seca nos meristemas de <i>C. canephora</i>	126
4	CONCLUSÃO.....	128
	REFERÊNCIAS.....	129

CAPÍTULO 1

INTRODUÇÃO GERAL

1 INTRODUÇÃO

O café é uma das *commodities* agrícolas mais importantes mundialmente e é o principal meio de subsistência para mais de 125 milhões de pessoas, sendo produzido em mais de 60 países. Enquanto que 126 espécies dos gêneros *Coffea L.* e *Psilanthus Hook. f.* são conhecidas atualmente (DAVIS et al., 2006), a produção de café comercial baseia-se principalmente em duas espécies relacionadas: *C. arabica L.* e *C. canephora Pierre*, cada uma com aproximadamente 65% e 35% da produção mundial respectivamente. O Brasil detém a liderança internacional em produção e exportação desse produto (INTERNATIONAL COFFEE ORGANIZATION- ICO, 2013).

De acordo com a Companhia Nacional de Abastecimento - CONAB (2013), a primeira estimativa da produção de café (arábica e conilon) para a safra 2013, indica que o país deverá colher entre 46,98 e 50,16 milhões de sacas de 60 quilos do produto beneficiado. A área total cultivada está estimada em 2,37 milhões de hectares no ano de 2013, enquanto que a área em produção totaliza em 2,066 milhões de hectares. A área em formação de café, ou seja, aquela que ainda não entrou no processo produtivo, apresenta crescimento de 10,49%.

A crescente demanda global por alimentos de melhor qualidade tem promovido um avanço significativo em diversas áreas da produção vegetal (NELSON et al., 2009). O café tem sido alvo de programas de melhoramento visando à introdução de novos caracteres para obtenção de híbridos com características agronômicas superiores, como floração uniforme, rendimento, tamanho do grão, qualidade da bebida, menor teor de cafeína e resistência às doenças e à seca.

Contudo, com o aquecimento global, espera-se um cenário de climas mais extremos, como secas, inundações e ondas de calor mais frequentes. Se

comprovados os cenários atuais preconizados pelos modelos do Painel Intergovernamental sobre Mudanças Climáticas (INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE - IPCC, 2013), onde está previsto um aumento na temperatura média anual do globo, o cultivo do café arábica nos estados de Minas Gerais, Goiás, São Paulo, Paraná, entre outros, será drasticamente reduzido nos próximos 100 anos, se mantidas as condições genéticas e fisiológicas das atuais variedades (ASSAD et al., 2004).

Por se tratar de uma cultura perene e de período juvenil longo, o melhoramento genético do cafeeiro é lento, sendo que a introdução de uma nova característica em uma variedade elite de café, por meio de técnicas de melhoramento convencional é um processo demorado (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 (ANTHONY et al., 2002). Além disso, a possibilidade de produção de plantas transgênicas abre novas perspectivas ao melhoramento convencional, permitindo a rápida incorporação de características desejáveis às espécies perenes (RIBAS; PEREIRA; VIEIRA, 2006).

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. Um número considerável dessas sequências gênicas está disponível nos bancos de dados do café, representando recursos valiosos para estabelecer um grande catálogo gênico para o gênero *Coffea* (MONDEGO et al., 2011).

Resultados de análises de bioinformática (MONDEGO et al., 2011; VIEIRA et al., 2006) mostram que aproximadamente 37% dessas sequências não possuem nenhuma similaridade, (sequências denominadas *no hits*) com aquelas conhecidas no GenBank. Estes *no hits* podem ser genes que desempenham papéis cruciais em determinados processos biológicos do cafeeiro, visto que a caracterização destes genes desconhecidos é um dos principais desafios da biologia moderna e o foco deste trabalho.

Proteínas desconhecidas podem estar envolvidas em vias biossintéticas importantes, na regulação gênica, nos eventos moleculares, fisiológicos e bioquímicos da planta de café, assim como na relação entre os estresses em que as plantas estão diariamente submetidas. Assim, o objetivo com este trabalho foi identificar e caracterizar funcionalmente, por meio de análises *in silico* e por qPCR da expressão gênica, *no hits* encontrados nas bases de dados do café.

2 REFERENCIAL TEÓRICO

2.1 O Café

2.1.1 Caracterização, Importância e Economia

Nativo do continente africano, o café chegou ao Brasil em 1723 por meio de mudas oriundas da Guiana Francesa. No ano seguinte, foi introduzido no Maranhão onde se propagou, para os Estados vizinhos, em pequenas plantações, tendo atingindo a Bahia em 1773. Algumas sementes de café foram transportadas do Maranhão para o Rio de Janeiro onde se fomentou a ampliação da cultura na Serra do mar (ALONSO-SALCES et al., 2009; CARVALHO et al., 2008).

Em 1825, as plantações alcançaram o Vale da Paraíba e os Estados de São Paulo e Minas Gerais. Posteriormente, a produção se estendeu para o centro-sul cujas condições ecológicas eram altamente favoráveis atingindo o norte do Rio de Janeiro e o Espírito Santo em 1920. A cadeia produtiva do café brasileiro se destaca na história econômica e social do país desde a época colonial, contudo, a exportação de café apresentou grande êxito até o ano de 1930, quando começou a decrescer em virtude do surgimento de crises internacionais (ALONSO-SALCES et al., 2009; CARVALHO et al., 2008).

Atualmente, o café é uma das *commodities* agrícolas mais importantes no mundo sendo responsável por quase metade do total das exportações de produtos tropicais além de ser o segundo maior, no mercado mundial de produtos naturais, depois do petróleo, e fonte de renda para muitos países em desenvolvimento (BRASIL, 2011).

Nos quatro primeiros meses do ano cafeeiro de 2012/13 (outubro/12 a janeiro/13) as exportações mundiais de café aumentaram 15,8% em relação às

exportações do mesmo período do ano cafeeiro passado, totalizando 37,87 milhões de sacas, ante 32,69 milhões. Nos 12 meses de fevereiro de 2012 a janeiro de 2013, as exportações de Arábica totalizaram quase 67,08 milhões de sacas, ante 66,11 milhões, e as de Robusta passaram para 47,59 milhões de 37,68 milhões de sacas (ICO, 2013).

A Associação Brasileira da Indústria de Café - ABIC (2012) estimou o aumento do consumo 4,0% em 2001, e vem crescendo sistematicamente desde 2003. As exportações de café estão projetadas para 2019/2020 em 37,7 milhões de sacas. Para obter essa estimativa, a taxa anual deverá se expandir em 1,7% (BRASIL, 2011). As duas espécies do gênero *Coffea* que são responsáveis por quase toda a produção dos grãos de café são *Coffea arabica* (Arábica) e *Coffea canephora* (Robusta), tem aproximadamente 65% e 35% da produção mundial (ICO, 2013), respectivamente, sendo que o café Arábica é considerado superior ao Robusta devido às suas propriedades organolépticas, e, portanto, possui um preço de mercado mais elevado.

O Brasil, o Vietnã e a Colômbia são responsáveis por mais de 50% da produção mundial desta cultura, sendo o Brasil com 35%, o maior produtor (ICO, 2013). Esta *commodity* é de primordial importância tanto para os produtores, que estão localizados nos países das áreas tropicais e subtropicais, e tem esta iguaria como seu principal produto de exportação agrícola; quanto para os fabricantes, que além de estarem amplamente envolvidos na industrialização e no comércio do café, são grandes consumidores da bebida. Estes estão localizados principalmente na Europa e América do Norte, onde o café é torrado, misturado, e embalado. Apesar de diversos fatores abióticos como seca e geada e, de fatores bióticos como pragas e doenças terem determinado oscilações na sua produção total, o Brasil ainda ocupa uma posição dominante no mercado mundial (CONAB, 2013; ICO, 2013).

O café é uma planta perene, pertencente à família *Rubiaceae*, subfamília *Ixoroideae*, tribo *Coffeae*, que é composta por dois gêneros: *Coffea L.* e *Psilanthus Hook f.*, estes juntos, abrangem mais de 100 espécies. O gênero *Coffea L.*, que teve seu centro de origem na África, contém as três espécies utilizadas na produção e bebida do café: *C. arabica*, *C. canephora* e *C. liberica* (DAVIS et al., 2006). *Coffea L.* é caracterizado por plantas dicotiledôneas, de folhas persistentes e flores hermafroditas, porte arbustivo ou arbóreo e caule lenhoso (FAZUOLI et al., 2000). Evolutivamente, o início da diversificação deste gênero ocorreu entre 5 e 25 milhões de anos atrás (CUBRY et al., 2008).

C. arabica é uma espécie alotetraploide autógama (anfidiplóide; $2n = 4 \times = 44$), e é originária de um cruzamento relativamente recente (aproximadamente 10000 a 50000 anos atrás) entre *C. canephora* e *C. eugenioides* (ou ecotipos relacionados a estas duas espécies), que ocorreu nos planaltos da Etiópia Central (CENCI; COMBES; LASHERMES, 2012; LASHERMES; BERTRAND; ETIENNE, 1999). *C. eugenioides* é uma espécie diplóide ($2n = 2 \times = 22$) selvagem que cresce em altitudes mais elevadas e produz poucos frutos e grãos pequenos e com baixo teor de cafeína (MAURIN et al., 2007).

Como consequência de sua história evolutiva e autogamia, as plantas de *C. arabica* têm uma estreita base genética, pois a difusão inicial a partir de seu centro de origem, a Etiópia, para o Iêmen, baseou-se em uma limitada população de plantas. Este problema é ampliado nos principais genótipos cultivados como, Mundo Novo, Catuaí e Caturra, que foram selecionados a partir de apenas duas populações base: Typica e Bourbon (ANTHONY et al., 2002). Por outro lado, *C. canephora* é um diplóide ($2n = 2 \times = 22$), alógama e uma espécie de café mais polimórfica. Devido ao modo de reprodução, a diversidade presente nesta espécie é mais ampla.

C. arabica, que é cultivado em ambientes de planalto é considerado como tendo uma melhor qualidade da bebida. Isto parece depender da qualidade e quantidade de compostos, tais como açúcares, armazenados no endosperma do fruto durante a maturação (GEROMEL et al., 2006; JOËT et al., 2009; LEROY et al., 2006; PRIVAT et al., 2008). *C. canephora* é mais bem adaptado às planícies equatoriais quentes e úmidas sendo considerado mais resistente a doenças e pragas tendo um maior teor de cafeína do que *C. arabica* (LEROY et al., 2005).

C. canephora tem sido caracterizado como um café neutro, de fraco sabor e, ocasionalmente, com um amargor forte e pronunciado, enquanto que o Arábica é mais leve, frutado, e de bebida acidulada (LEROY et al., 2006). Há também áreas propícias ao melhor cultivo de café, por exemplo, os Arábicas produzidos na América Central são tradicionalmente os mais apreciados, enquanto que o café arábica brasileiro ainda é considerado de qualidade inferior, devido aos métodos de colheita e práticas de processamento utilizado no país (AGRESTI et al., 2008).

2.1.2 Melhoramento genético do cafeeiro

Desde os primórdios da agricultura, as plantas vêm sendo melhoradas geneticamente pelo homem. Até recentemente, a única forma de se introduzir uma característica de interesse em um indivíduo ou espécie era por meio do melhoramento clássico, envolvendo cruzamentos e seleção da característica desejada. Essa seleção, em geral, era realizada apenas com base no fenótipo ou na avaliação de progênies (HERRERA et al., 2002).

No entanto, métodos convencionais de introdução de características desejáveis esbarram em uma série de limitações, como as barreiras de isolamento filogenético entre e dentro de grupos gênicos e a ligação gênica,

além do tempo necessário para transferir os caracteres de interesse (BORÉM; ALMEIDA, 2011).

Por se tratar de uma cultura perene e de período juvenil longo, o melhoramento genético do cafeeiro é lento. A introdução de uma nova característica em uma variedade elite por meio de técnicas de melhoramento convencional é um processo demorado (ANTHONY et al., 2002), podendo levar no mínimo 20 anos. Para os cultivares Obatã e Catuaí, por exemplo, o programa de melhoramento estendeu-se de 1968 a 2000 e de 1950 a 1992, respectivamente (CARVALHO et al., 2008). No entanto, por melhoramento clássico, já foram obtidos em *C. arabica* resistência à ferrugem do cafeeiro, a nematoide e ao fungo *Colletotrichum kahawae* (BERTRAND et al., 2003).

A diversidade genética de cultivares de *C. arabica* L. é relativamente pequena e a sua ampliação torna-se importante para o futuro do melhoramento do cafeeiro, visto que, trata-se de uma espécie nobre e que produz café de boa qualidade.

Esta estreita diversidade deve-se ao fato de serem plantas autógamas com centro de origem limitado e de terem sido utilizadas pequenas quantidades de sementes durante o processo de dispersão (CARVALHO; FAZUOLI, 1993), além de serem bastante aparentadas e em grande maioria, derivam-se das cultivares Típica, originalmente introduzida no Brasil em 1727 e Bourbon Vermelho, oriunda da ilha de mesmo nome (ANTHONY et al., 2001). Existem muitas variedades de *C. arabica* e mais de 40 mutantes já descritos (CARVALHO et al., 1991). Entretanto, devido ao modo de reprodução, a diversidade genética presente na espécie *C. canephora* é mais ampla que na espécie de *C. arabica*.

Montagnon, Cubry e Leroy (2012) analisaram a diversidade genética de plantas de *C. canephora* por meio de marcadores moleculares e mostraram a

existências de dois grupos distintos (Figura 1), estabelecidos em função da sua origem geográfica denominados: Guineano e Congolês

O grupo Guineano é constituído por populações selvagens da Costa do Marfim, na África ocidental, cujas principais características são: internódios, sementes e frutos pequenos com maturação precoce e altos teores de cafeína (em torno de 2,7%), na maior parte das plantas, suscetibilidade à ferrugem, tolerância à seca, bebida inferior ao grupo Congolês e coloração dos brotos novos frequentemente bronze (CUBRY et al., 2008; MONTAGNON; CUBRY; LEROY, 2012).

Já o grupo Congolês apresenta dois subgrupos de maior importância: o subgrupo 1-(SG1) (tolerante à seca), formado pelos tipos de café Robusta ou Híbridos entre os dois grupos (Kouillou x Robusta) e o subgrupo 2-(SG2) sensível à seca que corresponde ao café Robusta propriamente dito (MONTAGNON; LEROY, 1993).

Os representantes do grupo Congolês apresentam internódios longos, frutos grandes, maturação média a tardia dos frutos, sementes grandes, peso das sementes maiores do que as do grupo Guineano, peneira média alta, menores teores de cafeína (em torno de 2,5%), maior resistência à ferrugem, suscetibilidade à seca, exigência em água, bebida superior ao Grupo Guineano e coloração das folhas novas bronze ou marrom (CUBRY et al., 2008; MONTAGNON; CUBRY; LEROY, 2012).

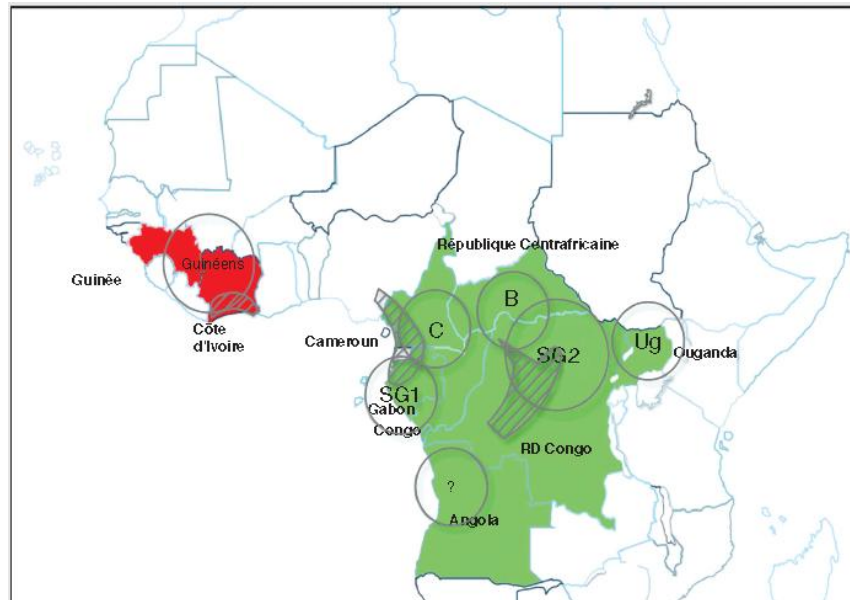


Figura 1 Origem geográfica dos principais grupos genéticos de *C. canephora*. Vermelho: origem geográfica do grupo Guineano, verde: origem geográfica do grupo genético congolês; círculos com o nome dos grupos genéticos: Origem geográfica de cada grupo; áreas hachuradas: refúgios do último máximo glacial (20.000 anos atrás). Fonte: Montagnon, Cubry e Leroy (2012)

Recentemente, métodos avançados de melhoramento genético vegetal têm se destacado, principalmente, devido à possibilidade de obtenção de resultados uniformes e direcionados para características desejáveis. Dentre as técnicas de biotecnologia moderna que poderiam auxiliar o programa de melhoramento do cafeeiro destacam-se a cultura de tecidos, a transformação genética e a seleção assistida por marcadores moleculares. O melhoramento genético de plantas via engenharia genética tornou-se uma ferramenta importante na obtenção de plantas com novas características agrônômicas em um curto espaço de tempo, sendo, portanto, de grande interesse para o setor agrícola (BORÉM, 1998).

O uso eficiente dos recursos disponíveis na genética de espécies selvagens diploides é essencial para a melhoria contínua das variedades de café arábica (LASHERMES et al., 2009; VOSSEN et al., 2001). Novos *insights* sobre a introgressão em *C. arabica* são particularmente valiosos para melhorar a metodologia de melhoramento atual (LASHERMES et al., 2011; MARRACCINI et al., 2011; VIDAL et al., 2010).

Estudos mostraram a existência de regulações gênicas específicas em *C. arabica* devido a presença dos genes homeólogos provenientes dos sub-genomas de *C. canephora* (*CaCc*) e *C. eugenioides* (*CaCe*). Os homeólogos de *CaCe* parecem ser genes que se expressam para assumir as funções básicas das células, enquanto que *CaCc* parecem ser genes implicados na regulação (VIDAL et al., 2010).

Outro desafio ao melhoramento genético é o representado pelo controle de pragas e doenças, bem como o aumento do rendimento e adaptabilidade endofoclimática das espécies agrônomicas, sendo que uma das estratégias fundamentais de melhoramento para lidar com o desafio das mudanças climáticas é o de aprimorar a adaptação das culturas para que seja possível um período de safra curto que ligue a fenologia à disponibilidade de umidade (CECCARELLI et al., 2010).

A introdução direta de genes por engenharia genética parece uma solução atraente e rápida para melhorar a tolerância a diversos estresses (DUNWELL, 2000; RIBAS; PEREIRA; VIEIRA, 2006; WANG; VINO CUR; ALTMAN, 2003). Esta tem sido aplicada com sucesso para combater dezenas de pragas e para a redução de ervas daninhas. Assim, decifrar os mecanismos pelos quais as plantas percebem sinais ambientais e sua transmissão para a maquinaria celular para ativar respostas adaptativas é de fundamental importância para o desenvolvimento do melhoramento racional e estratégias de transgenia (AGARWAL et al., 2006). Melhorias nas características agrônomicas de café,

como a qualidade da bebida, resistência a patógenos e insetos e tolerância ao estresse hídrico são muito importantes para o setor cafeeiro.

Entretanto, os genes são os responsáveis por ditar essas características inerentes às espécies e eles são o foco dos estudos em genética, sendo estes estudos a base para a aplicação prática do melhoramento genético nas espécies. É extremamente importante conhecer a função dos genes, seja ela a nível molecular, celular, de organismo, familiar, populacional ou evolutivo. Sendo o foco deste estudo, a caracterização e análise dos genes que ainda não foram descritos e analisados.

2.1.3 Estresse Biótico e Abiótico

As plantas respondem e se adaptam as condições de estresses com uma série de alterações morfológicas, fisiológicas e moleculares. No nível molecular, a expressão diferencial de genes para tolerar as condições adversas é um bom indicativo da resposta da planta ao meio ambiente. Encontrar a função destes genes facilitará o entendimento da relação planta-ambiente e planta-patógeno.

No início do século XX, o Brasil teve uma acentuada redução da área cafeeira devido a diversos fatores que têm determinado oscilações na sua produção total, dentre estes, fatores abióticos como seca, alto teor de sal e mudanças de temperatura e bióticos como pragas e doenças (GONÇALVES, 1999).

O estresse ambiental apresenta um grande desafio na busca de uma produção sustentável de alimentos, pois reduz os rendimentos potenciais em até 70% das plantas cultivadas (AGARWAL et al., 2006). O estresse hídrico ocasionado pela seca e severidade de temperatura é o estresse abiótico mais prevalente, que limita o crescimento das plantas e produtividade. O fechamento dos estômatos é uma das respostas à secagem do solo,

consequentemente limitando a perda de água e a assimilação do carbono pela fotossíntese (FLEXAS; MEDRANO, 2002).

Múltiplas vias de sinalização regulam as respostas de estresse nas plantas (KNIGHT; KNIGHT, 2001) e existe uma sobreposição entre os padrões de expressão de genes que são induzidos em resposta a diferentes fatores de estresse (CHEN et al., 2002; SEKI et al., 2001). Estresses abióticos são, na verdade estímulos complexos que induzem muitos atributos diferentes, mas relacionados, o desequilíbrio iônico e o estresse osmótico, que podem fornecer às células informações exclusivas.

Com base nesta multiplicidade de sinalização, pode haver vários sensores que percebem o sinal de estresse inicial e alteraram a expressão de um grande número de genes. Respostas moleculares e celulares aos estresses abióticos e bióticos incluem a percepção, transdução de sinal para o citoplasma e núcleo, expressão do gene, e, finalmente, alterações metabólicas levando à tolerância ao estresse (AGARWAL et al., 2006).

Análises fisiológicas sugerem que a tolerância à seca pode ser uma consequência direta de um melhor desenvolvimento da raiz (PINHEIRO et al., 2005). Em outro estudo, Lima et al. (2002) sugeriram que a atividade aumentada de enzimas antioxidantes também pode estar envolvida no mecanismo de tolerância à seca.

Geralmente, além da seca, as plantas no campo enfrentam inúmeros estresses coletivamente, sejam eles bióticos ou abióticos. Alguns relatórios recentes puseram em evidência a ligação entre a resistência a doenças e tolerância à seca (AGARWAL et al., 2006).

Para estresses abióticos, a engenharia das “proteínas de estresse” ou as enzimas das vias de biossíntese associadas com as respostas ao estresse tem sido desenvolvida como um método de incentivo para melhorar a tolerância ao estresse. Loon, Rep e Pieterse (2006) mostram em seu estudo que proteínas PRs

(“*Pathogenesis Related proteins*”) são um grupo heterogêneo de proteínas vegetais, induzidas por estresses bióticos. Algumas destas proteínas são efetores contra patógenos e insetos, enquanto que outras estão envolvidas no restabelecimento da homeostase após o estresse.

2.1.4 Desenvolvimento do Fruto

O sabor e aroma do café são os produtos de uma complexa cadeia de transformações químicas. O grão verde tem apenas um cheiro fraco, mas contém todos os precursores necessários para gerar o aroma do café durante a torrefação. Um dos principais objetivos da pesquisa sobre a qualidade do café é o de estabelecer relações entre a composição química do grão verde de café e a qualidade da bebida, e em particular, para identificar os precursores gênicos envolvidos no desenvolvimento do aroma e sabor.

No entanto, diferenças entre *C. arabica* e *C. canephora* têm sido relatadas com relação à maturação dos frutos, onde a maturação dos frutos do café Robusta é atrasada em comparação com o Arábica (LEROY et al., 2006).

O processo de desenvolvimento de frutos compreende as etapas de crescimento, maturação, amadurecimento e senescência (PEZZOPANE et al., 2003). A fase de crescimento é responsável por um aumento de atributos físicos de um fruto em desenvolvimento. A maturação corresponde ao estágio do desenvolvimento que leva à maturidade fisiológica. O amadurecimento é onde ocorre uma série de processos no final do desenvolvimento e que resultam em características estéticas e de qualidade, evidenciadas por mudanças na composição, coloração, textura, sabor e aroma. A senescência se caracteriza por uma série de processos que ocorrem após a maturação e amadurecimento, levando à morte dos tecidos (CASTRO; KLUGE; PERES, 2005).

No café o desenvolvimento dos frutos é um processo longo, caracterizado por mudanças e evoluções nos tecidos (CASTRO; MARRACCINI, 2006). Logo após a fecundação e até a fase de maturação, o fruto é principalmente constituído pelo pericarpo e perisperma. Em seguida, o perisperma vai desaparecendo, sendo substituído progressivamente pelo endosperma que endurece durante a fase de maturação devido ao acúmulo gradual de proteínas de reserva, sacarose e polissacarídeos complexos, representando as principais reservas da semente. Na maturação, ocorre desidratação do endosperma e alteração da cor do pericarpo (LEROY et al., 2006).

O pericarpo é composto por vários tecidos como exocarpo, mesocarpo e endocarpo. O exocarpo é responsável pela expressão da coloração do fruto do café durante o desenvolvimento. O mesocarpo é rico em açúcar e água, sendo formado de células do parênquima com paredes celulares compactas e densas em frutos verdes, que durante a maturação se tornam mais finas provavelmente devido a modificações de pectina. O endosperma é localizado na parte mais interna do pericarpo (CASTRO; MARRACCINI, 2006). O perisperma desenvolve-se do núcleo do óvulo logo depois da fecundação e permanece até a fase madura (DENTAN, 1985). O endosperma de café é um tecido triplóide, assim como em outras espécies de plantas.

Estudos realizados por Lin et al. (2005), relacionados aos genes da via biossintética da cafeína, evidenciaram a participação dos mesmos em diferentes tecidos das plantas como endosperma e folhas jovens. Os genes da teobromina sintase e cafeína sintase foram expressos em frutos imaturos e a expressão do gene teobromina sintase foi mais alta, não somente em frutos verdes pequenos (fase chumbinho), mas também em pericarpo e em endosperma imaturo de frutos verdes maiores (fase verde). Os pesquisadores observaram por meio da tecnologia RT-PCR que não houve expressão de metilxantosina, teobromina e

cafeína sintase em frutos maduros, o que poderia explicar a ausência do acúmulo de cafeína durante o processo de maturação nos frutos avaliados.

Os perfis fenólicos e de metilxantina dos grãos de café verde são afetados por vários fatores: a variedade de café, as propriedades genéticas dos cultivares, a maturidade dos grãos na colheita, o método de colheita e as condições pós-colheita de processamento (fermentação, lavagem, secagem, armazenamento), práticas agrícolas (sombra, poda de fertilização), fatores ambientais (solo, altitude exposição ao sol), e os parâmetros climáticos (chuva, temperatura).

Castro e Marraccini (2006) mostraram a caracterização de perfis de expressão gênica durante as mudanças fenológicas que ocorrem durante o desenvolvimento da semente, o que permite a identificação dos principais genes relacionados com a qualidade. No entanto, poucos estudos têm procurado identificar os genes envolvidos no determinismo molecular da composição bioquímica dos grãos de café.

Segundo Salmona et al. (2008), os frutos de café arábica possuem os seguintes estádios de desenvolvimento (Figura 2): Estádio 1 (00-60 DAF, dias após a floração): sementes, principalmente formada por perisperma aquoso (frutos pequenos com peso de 0.5 g); Estádio 2 (60-90 DAF): muito semelhante ao estádio 1, exceto que o fruto de peso é superior a 0.5 g. A semente é formada principalmente de perisperma aquosa em torno de um endosperma líquido muito pequeno, Estádio 3 (90-120 DAF): tecido de endosperma aquoso crescente e progressivamente substituindo o perisperma no lóculo, Estádio 4 (120-150 DAF): endosperma leitoso macio, Estádio 5 (150-210 DAF): endosperma duro branco, enquanto que o perisperma restante se assemelha a uma película fina verde envolvendo o endosperma, Estádio 6 (210-240 DAF): amadurecimento,

frutos cereja com pericarpo tendendo para o amarelo, Estádio 7 (240 DAF):
frutos maduros vermelhos cereja com pericarpo transformando-se em roxo.

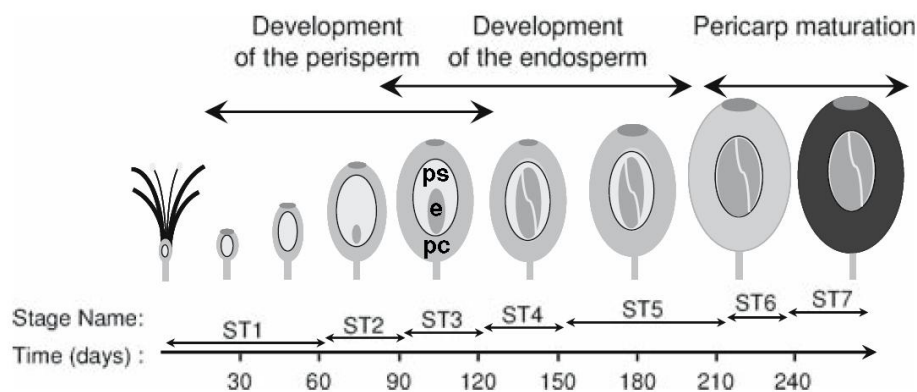


Figura 2 Gráfico das alterações teciduais que ocorrem durante o desenvolvimento dos frutos de café e maturação. e: endosperma; ps: perisperma; pc: pericarpo. Fonte: Salmona et al. (2008), adaptado de Castro e Marraccini (2006).

2.2 Bioinformática

2.2.1 Estudo de genomas expressos (ESTs) em plantas

Como o advento das novas tecnologias de sequenciamento (ZHANG et al., 2011) e construção de projetos genoma, que geram uma quantidade maciça de dados que precisam ser organizados, arquivados e analisados; os métodos avançados em bioinformática são requeridos a cada dia para executar tais funções.

O sequenciamento completo de espécies com genomas grandes é demorado, oneroso e trabalhoso, apesar das técnicas modernas. O sequenciamento parcial de clones anônimos de cDNA, os ESTs (Expressed

Sequences Tags), são uma alternativa simples e eficiente para identificar e analisar genes expressos a partir de transcritos gênicos como descrito por Nagaraj, Gasser e Ranganathan (2007), Figura 3.

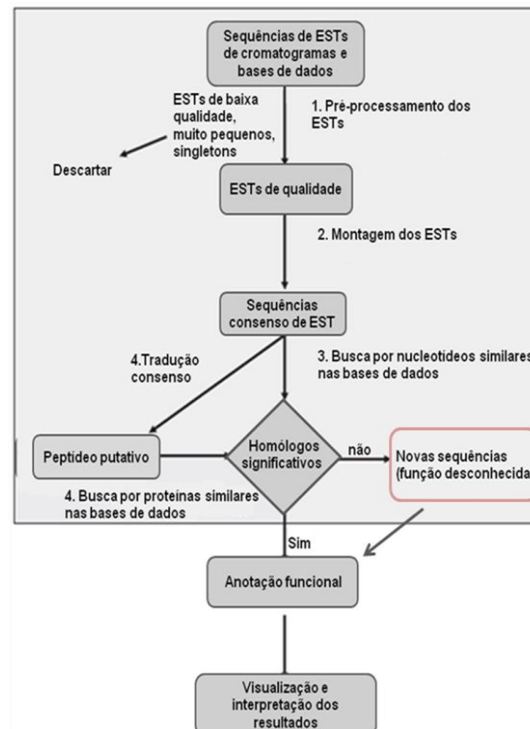


Figura 3 Passos gerais envolvidos na análise de ESTs. 1. Sequências primárias de ESTs são verificadas quanto à presença do vetor, regiões repetitivas, que são excisadas ou mascaradas. Sequências com baixa qualidade, *singletons* e sequências muito curtas também são removidas. 2. ESTs são então agrupadas e montadas para gerar sequências consenso (“transcritos putativos”). 3. Buscas por similaridade em banco de dados de DNA são realizadas para identificar homólogos e atribuir uma possível função. 4. Putativos peptídeos são obtidos por tradução das sequências consenso. 5. Buscas por similaridade no banco de dados de proteínas são realizadas para atribuir a(s) função (ões) putativa(s). A análise é estendida pela anotação funcional seguida de visualização e interpretação dos resultados. Adaptado de Nagaraj, Gasser e Ranganathan (2007).

A estratégia básica de obtenção de EST constitui um artifício eficiente de amostragem do genoma para sequências ativas de genes. Eles representam uma sequência total ou parcial de um gene que foi expresso em um tecido, em um momento determinado (STERKY; LUNDEBERG, 2000). Atualmente, os ESTs permitem a descoberta de genes, anotação do genoma complemento, ajuda na identificação da estrutura do gene, a estabelecer a viabilidade de transcritos alternativos, orientam polimorfismos (SNP) e facilita a análise do proteoma (NAGARAJ; GASSER; RANGANATHAN, 2007).

A análise de ESTs se tornou a ferramenta que mais cresce nas bases de dados públicas de DNA. Por esta razão, vários projetos de sequenciamento de espécies de plantas, como o Projeto Genoma de ESTs de cana-de-açúcar (SUCEST) realizado pelo grupo ONSA (Organização para Análise e Sequenciamento de Nucleotídeos) (GRIVET; ARRUDA, 2002) têm sido refeitos no Brasil.

Nas plantas, ESTs foram inicialmente utilizadas para a espécie modelo *Arabidopsis thaliana* (HÖFTE et al., 1994) e arroz (YAMAMOTO; SASAKI, 1997). Posteriormente, uma grande variedade de sequências de outras espécies tem sido depositadas na dbEST (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION - NCBI, 2012).

Graças à exploração de informações de expressão gênica nas espécies modelos, por meio dos ESTs, aliada à conservação gênica entre espécies relacionadas (também conhecido como sintenia), foi possível a identificação de genes candidatos associados à características agronômicas relevantes. Como exemplo, Marraccini et al. (2012) identificaram 13 genes candidatos em café, por meio de análises de bioinformática, que respondem diferencialmente ao estresse hídrico. Guzzo, Harakava e Tsai (2009) identificaram em café genes envolvidos no mecanismo de defesa descrito em outras espécies de plantas por meio da busca nos bancos de dados de ESTs. Mondego et al. (2005) isolaram e

caracterizaram genes envolvidos na resistência ao bicho-mineiro em café, selecionando ESTs expressos nas bibliotecas de café resistentes à *Leucoptera coffeella*.

Outro exemplo recente foi o da identificação de genes candidatos por meio do sequenciamento em larga escala (454-GS-FLEX Titanium) de cDNAs de folhas de café infectadas com ferrugem, onde foram identificados por meio da busca por similaridade, sequências de ESTs do fungo *Hemileia vastatrix* que não mostraram homologia com qualquer base de dados do genoma, indicando novos genes potenciais do fungo (FERNANDEZ et al., 2012). Vários estudos já utilizam também a estratégia de busca de SSR em Bancos de ESTs, como descrito por Cato et al. (2001).

É possível realizar análises de expressão quantitativa, correlacionando a frequência dos ESTs com características desejáveis das espécies de plantas. Constitui também uma ferramenta interessante para a detecção de genes responsivos a estresses e variação genética que pode explicar características fenotípicas específicas.

Além disso, estas análises podem fornecer metas para transgenia e *insights* sobre o desenvolvimento das plantas e culturas. Fornecendo assim uma fonte para a descoberta de novos genes e para análises comparativas entre os organismos como relatado em Mondego et al. (2011).

Neste estudo os pesquisadores verificaram a disponibilidade de dados de ESTs de ambas as espécies de *Coffea* mais comercialmente importantes, o que os levou a realizar uma ampla análise em bioinformática. Dentre estas análises cinco famílias gênicas foram específicas ou proeminentes em ambas *C. arabica* e *C. canephora* quando comparado com as outras espécies analisadas e foi observado que duas destas famílias contêm proteínas desconhecidas.

2.2.2 Bases de dados do café

O Projeto Brasileiro do Genoma Café - PBGC (MONDEGO et al., 2011; VIEIRA et al., 2006) foi desenvolvido para acrescentar e implantar ferramentas úteis para a descoberta de genes e análises genéticas funcionais em espécies de café e relacionados, e para auxiliar no avanço do conhecimento sobre a estrutura e evolução do genoma do café. Este projeto resultou na identificação de cerca de 33.000 unigenes diferentes vindos de 37 bibliotecas, correspondentes aos genes expressos nos vários tecidos da planta; folhas, raízes, frutos, flores e ramos; sadios e submetidos a estresses bióticos e abióticos como pragas, doenças, frio, calor, seca e em diversos estádios de desenvolvimento (Tabela 1).

Tabela 1 Descrição das bibliotecas de ESTs geradas pelo Projeto Brasileiro do Genoma Café. Fonte: Vieira et al. (2006)

Bibliotecas	Tecido/Estádios de desenvolvimento	Números de reads válidos
AR1, LP1	Plântulas e folhas tratadas com ácido araquidônico	5664
BP1	Suspensão celular tratada com acibenzolar-S-metil	12379
CB1	Suspensão celular tratada com acibenzolar-S-metil e brassinosteroides	10311
CL2	Hipocótilos tratados com acibenzolar-S-metil	11615
CS1	Suspensão celular tratada com NaCl	10803
EA1, IA1, IA2	Calos embriogênicos	9191
EB1	Embrião zigótico (frutos imaturos)	192
EC1	Calos embriogênicos de <i>Coffea canephora</i>	8050
EM1, SI3	Sementes em germinação (sementes inteiras e embriões zigóticos)	9201
FB1, FB2, FB4	Botões florais em diferentes estádios de desenvolvimento	23036
FR1, FR2	Botões florais + chumbinho + frutos em diferentes estádios	14779
FR4	Frutos (<i>Coffea racemosa</i>)	7967
FV2	Frutos, nos estádios 1,2 e 3 (<i>Coffea racemosa</i>)	7195
CA1, IC1, PC1	Calo não embriogênico com e sem 2,4 D	12135
LV4, LV5	Folhas jovens dos ramos ortotrópicos	15067
LV8, LV9	Folhas adultas dos ramos plagiotrópicos	11864
NS1	Raízes infectadas com nematoide	569

“Tabela 1, conclusão”

PA1	Calo embriogênico primário	2483
RM1	Folhas infectadas com bicho-mineiro e ferrugem	5567
RT3	Raízes	560
RT5	Raízes com acibenzolar-S-metil	2311
RT8	Suspensão celular estressadas com alumínio	9119
RX1	Ramos infectados com <i>Xylella spp.</i>	9563
SH1	Folhas de plantas em estresse hídrico (<i>Coffea canephora</i>)	7368
SH2	Plantas em estresse hídrico do campo (vários tecidos)	6824
SS1	Plantas irrigadas no campo (vários tecidos)	960

Os dados obtidos no PBGC foram armazenados em duas plataformas de bioinformática. Uma delas desenvolvida pelo Laboratório de Genômica e Expressão (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA, 2012b) e está disponível desde o segundo semestre de 2004. A outra plataforma, disponibilizada em março de 2007, foi criada pela Embrapa Recursos Genéticos e Biotecnologia – Cenargen (EMBRAPA, 2012a).

Sales, Barbosa e Silva (2007) apresentaram as ferramentas já existentes na base de dados hospedada na Embrapa Recursos Genéticos e Biotecnologia a permitir que os usuários do projeto consigam identificar rapidamente genes de interesse neste conjunto de dados.

Em geral, a bioinformática nos projetos de EST, como no caso do PGBC, inclui serviços tais como a organização, armazenamento, integração e análise da informação biológica desempenhando assim um papel crucial. Em particular, os padrões de qualidade de sequência e anotação têm que ser uniforme em todos os diferentes cromossomos gerados por meio dos diferentes projetos, e o acesso a dados eficientes tem de ser prestado à comunidade científica. Os objetivos com a bioinformática no Projeto Brasileiro do Genoma Café foram de estabelecer os meios em que vários tipos diferentes de dados possam ser armazenados e de fácil manipulação, para atender às necessidades dos usuários (VIEIRA et al., 2006).

No entanto, outros bancos de dados de café estão disponíveis, como no caso da iniciativa da Nestlé Research Center e do Departamento de Biologia

Vegetal da Universidade de Cornell, um banco de dados EST com base em sequências de aproximadamente 47.000 clones de cDNA e com um foco especial em sementes em desenvolvimento de *C. canephora* também foi feito (LIN et al., 2005; MUELLER et al., 2005). Neste caso foram sequenciados ESTs provenientes de *C. canephora* crescidos no campo no leste de Java, na Indonésia, gerando um total de 13.175 unigenes. Este grupo de pesquisa comparou também os repertórios de EST de *C. canephora*, *Solanum lycopersicum* (tomate) e *A. thaliana*.

O grupo de pesquisa CENICAFE, por sua vez, produziu sequências de EST de três tecidos diferentes: folhas, frutos e flores de *C. arabica* (cv. Caturra) montadas em 10.799 unigenes (MONTROYA et al., 2006). Outro grupo Francês, o IRD, produziu 5.534 unigenes potenciais a partir das bibliotecas de cDNA de frutos e folhas (PONCET et al., 2006). 55.694 sequências tornaram-se disponíveis formando uma das principais fontes de pesquisa para a comunidade científica.

Com um número exuberante de informações disponíveis presentes nas bases de dados do café, será possível esclarecer de forma rápida e eficiente os mecanismos de regulação gênica, de evolução, de sinalização, de resistência à doenças, de defesa, de desenvolvimento dos frutos e de qualidade dos mesmos, e, finalmente, contribuir para a melhoria da agricultura por meio da diversidade fenotípica encontrada.

2.3 NO HITS

Entre as diferentes ferramentas em bioinformática, uma das mais relevantes na anotação de ESTs recém sequenciados é o Blast. O Blast permite determinar a possível função de um gene, pela busca de similaridade com sequências de genes conhecidos descritos em bancos de dados públicos (ex:

NCBI). A atribuição de uma provável função a uma sequência EST só é possível graças à existência de genes ortólogos (genes homólogos, descritos em diferentes espécies, que se originam de um gene ancestral comum por especiação). Dessa forma, pode-se atribuir uma possível função a uma sequência EST pelo grau de similaridade entre sequências provenientes de organismos distintos. Uma característica comum de genes ortólogos é de que eles realizam a mesma função (KOONIN, 2005).

Algumas vezes, mesmo fazendo todas essas análises, não é fácil encontrar alguma informação que permita associar a sequência EST em estudo com algum gene já descrito ou com alguma proteína. De fato, nos bancos de dados são encontrados termos para designar uma proteína desconhecida, como: *Hypothetical protein*, *Putative protein*, *Predicted protein* e *Protein with unknown function*.

No entanto, apesar de não haver uma definição clara desses termos, assume-se *Hypothetical protein* como a uma sequência de aminoácidos que não se tem certeza que origina uma proteína; *Putative* e *Predicted protein* correspondem a proteínas cuja função necessita de comprovação experimental, e *Protein with unknown function* corresponderia a uma sequência de aminoácidos que comprovadamente se constitui em uma proteína, no entanto sua função é desconhecida (HOULLOU-KIDO et al., 2010).

Dessa forma, há sequências que não possuem nenhuma similaridade com aquelas conhecidas nos bancos de dados quando é feita uma busca por homologia por meio da ferramenta BLAST. Estas sequências são denominadas neste trabalho de *no hits* (usando a ferramenta BLASTp, filtrados por meio de um de E-value de $\leq 1e^{-1}$). Estas sequências são tratadas como órfãs (TAUTZ; DOMAZET-LOSO, 2011) e representam parcelas significativas nas bases de dados. Estes *no hits* podem ser genes que desempenham papéis cruciais em determinados processos biológicos do cafeeiro, visto que a caracterização destes

genes desconhecidos é indispensável. Proteínas desconhecidas podem estar envolvidas em vias biossintéticas importantes, na regulação gênica, nos eventos moleculares, fisiológicos e bioquímicos do café, assim como na relação entre os estresses em que as plantas estão diariamente submetidas.

Outros ácidos ribonucleicos não codificadores de proteínas (miRNAs) são um importante conjunto de RNAs dentro da totalidade dos transcritos existentes nas células e estes podem corresponder a alguns *no hits*. Os miRNA contêm nas suas moléculas informações variadas que lhes permitem ter diversas funções (BARTEL, 2007).

Tautz e Domazet-Loaso (2011) discutem sobre a origem evolutiva, a permanência e a seleção adaptativa destes genes órfãos (filtrados por meio de um de E-value de $\leq 1e^{-4}$) em determinadas linhagens no decorrer da evolução.

Alguns *no hits* estão sendo identificados e caracterizados como nos estudos recentes de Marraccini et al. (2012), que relataram genes candidatos potencialmente envolvidos na tolerância à seca em cafeeiro, por meio de análises da expressão gênica, utilizando clones tolerantes e sensíveis de café robusta (*C. canephora* var. *conillon*) (Figura 4). Dentre eles, os *no hits* *CcUNK01*, *CcUNK08* e *CcUNK10* apresentaram perfis de expressão diferencial entre os dois materiais genéticos testados, sendo que o *CcUNK10* e *CcUNK01* aumentaram sua expressão na condição de estresse hídrico, de maneira mais acentuada no clone sensível (22), quando comparado aos clones tolerantes (14, 120) analisados.

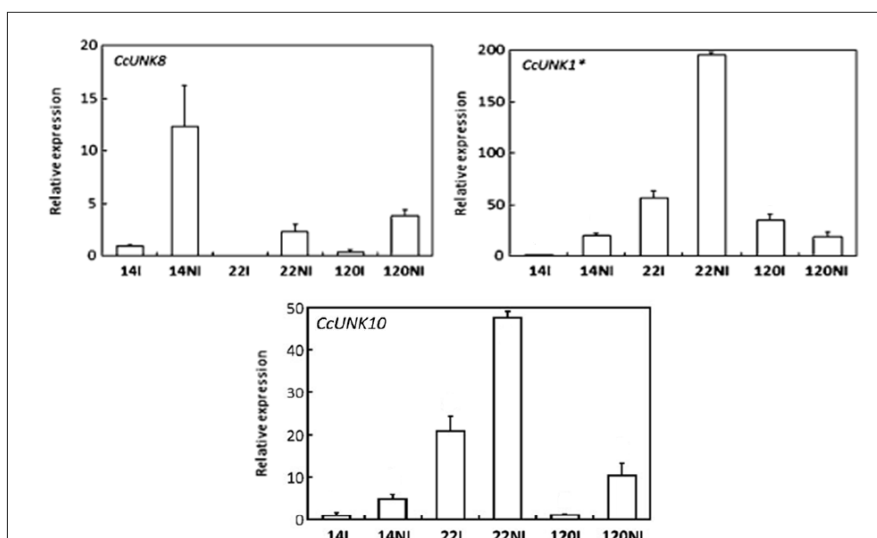


Figura 4 Perfis de expressão dos genes *CcUNK8*, *CcUNK1* e *CcUNK10* nas folhas dos clones 14, 22 e 120 de *C. canephora* submetidos (NI: Não irrigado) ou não (I: Irrigado) à seca. Resultados da expressão usando o 14I como calibrador interno. A expressão do gene *CcUBQ10* foi utilizada como referência para a quantificação relativa, exceto para o gene marcado com asterisco que foi usado o gene *CcGAPDH* como referência. Adaptado de Marraccini et al. (2012).

Horan et al. (2008) identificaram e caracterizaram os genes que codificam proteínas de função desconhecida no genoma de *A. thaliana*. Outros trabalhos também têm identificado *no hits*, como em Alvarenga et al. (2010) que identificaram nas bases de dados do PGBC *no hits* potencialmente relacionados à resistência às doenças em café.

Já nos estudos em trigo, cerca de 50% dos ESTs estudados foram significativamente homólogos a genes que codificam para proteínas com funções conhecidas, 20% foram semelhantes aos genes que codificam proteínas com funções desconhecidas, e 30% não têm similaridade significativa com qualquer sequência nas bases de dados. Estes dados mostram ainda um número exuberante de informações que precisam ser catalogadas e estudadas (MA et al., (2009).

3 CONSIDERAÇÕES GERAIS

A identificação e caracterização funcional dos genes desconhecidos ou órfãos, chamados aqui de *no hits* podem oferecer recursos valiosos para o avanço das pesquisas científicas referentes ao café, auxiliando na elucidação de mecanismos moleculares, assim como vias metabólicas e processos biológicos e fisiológicos em que estes genes estão implicados. Fatores genéticos determinantes relacionados a diferentes características fenotípicas de interesse como, floração uniforme, rendimento, tamanho do grão, qualidade da bebida, teor de cafeína e resistência às doenças e à seca, são de interesse particular para a geração de novas variedades melhor adaptadas. Neste contexto, há uma demanda pela caracterização funcional de transcriptomas devido ao crescimento exponencial do número de novas seqüências gênicas descobertas nos projetos genoma ESTs como no Projeto Genoma Café e mais recentemente, vários Projetos de RNAseq.

REFERÊNCIAS

AGARWAL, P. K. et al. Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. **Plant Cell Reports**, Berlin, v. 25, n. 12, p. 1263-1274, Dec. 2006.

AGRESTI, P. et al. Discrimination between defective and non-defective Brazilian coffee beans by their volatile profile. **Food Chemistry**, London, v. 106, n. 2, p. 787-796, Jan. 2008.

ALONSO-SALCES, R. M. et al. Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): chemometric evaluation of phenolic and methylxanthine contents. **Journal of Agricultural and Food Chemistry**, Easton, v. 57, n. 10, p. 4224-4235, May 2009.

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

ANTHONY, F. et al. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica**, Wageningen, v. 118, n. 1, p. 53-65, June 2001.

_____. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. **Theoretical and Applied Genetics**, Berlin, v. 104, n. 5, p. 894-900, Apr. 2002.

ASSAD, E. D. et al. Climatic changes impact in agroclimatic zoning of coffee in Brazil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 11, p. 1057-1064, nov. 2004.

ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. **Tendências de consumo de café VIII, 2011**. Disponível em: <<http://www.abic.com.br/publique/cgi/cgilua.exe/sys/start.htm?sid=45>>. Acesso em: 20 jan. 2012.

BARTEL, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, Cambridge, v. 131, n. 4, p. 11-29, Nov. 2007.

BASIC LOCAL ALIGNMENT SEARCH TOOL. **GenBank**. Disponível em: <<http://blast.ncbi.nlm.nih.gov/Blast.cgi>>. Acesso em: 10 dez. 2012.

BERTRAND, B. et al. Impact of the *Coffea canephora* gene introgression on beverage quality of *C. arabica*. **Theoretical and Applied Genetics**, Berlin, v. 107, n. 3, p. 387-394, Aug. 2003.

BORÉM, A. **Biotecnologia no melhoramento de plantas**: melhoramento de Plantas. 2. ed. Viçosa, MG: UFV, 1998. 453 p.

BORÉM, A.; ALMEIDA, G. D. **Plantas Geneticamente modificadas**: desafios e oportunidades para regiões tropicais. Viçosa, MG: UFV, 2011. 390 p.

BRASIL. Ministério da Agricultura Pecuária e Abastecimento. **Projeções do agronegócio Brasil 2010/2011 e 2020/2021**. Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 17 out. 2011.

CARVALHO, A. et al. Genetics aspects of the coffee tree. **Revista Brasileira de Genética**, Ribeirão Preto, v. 14, n. 1, p. 135-183, Mar. 1991.

CARVALHO, A.; FAZUOLI, L. C. Café. In: FURLANI, A. M. C.; VIÉGAS, G. A. (Ed.). **O melhoramento de plantas no Instituto Agronômico**. Campinas: Instituto Agronômico, 1993. p. 29-76.

CARVALHO, C. H. S. et al. Cultivares de café arábica de porte baixo. In: _____. **Cultivares de café**: origem, características e recomendações. Brasília: EMBRAPA Café, 2008. p. 157-226.

CASTRO, P. R. C.; KLUGE, R. A.; PERES, L. E. P. **Manual de fisiologia vegetal**: teoria e prática. Piracicaba: Agronômica Ceres, 2005. 650 p.

CASTRO, R. D. de; MARRACCINI, P. Cytology, biochemistry and molecular changes during coffee fruit development. **Brazilian Journal of Plant Physiology**, Piracicaba, v. 18, n. 1, p. 175-199, Mar. 2006.

CATO, S. A. et al. A rapid PCR-based method for genetically mapping ESTs. **Theoretical and Applied Genetics**, Berlin, v. 102, n. 2/3, p. 296-306, Feb. 2001.

CECCARELLI, S. et al. Plant breeding and climate changes. **Journal of Agricultural Science**, Cambridge, v. 148, n. 1, p. 627-637, Aug. 2010.

CENCI, A.; COMBES, M. C.; LASHERMES, P. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. **Plant Molecular Biology**, Berlin, v. 78, n. 1/2, p. 135-145, Jan. 2012.

CHEN, W. Q. et al. Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. **Plant Cell**, Rockville, v. 14, n. 3, p. 559-574, Mar. 2002.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira: café**. Disponível em: <[http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_01_10_10_54_22_boletim_cafe_1a_estimativa\[1\].pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_01_10_10_54_22_boletim_cafe_1a_estimativa[1].pdf)>. Acesso em: 23 jan. 2013.

CUBRY, P. et al. Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. **Genome**, Ottawa, v. 51, n. 1, p. 50-63, Jan. 2008.

DAVIS, A. P. et al. An annotated taxonomic conspectus of the genus *Coffea* (*Rubiaceae*). **Botanical Journal of the Linnean Society**, London, v. 152, n. 4, p. 465-512, Dec. 2006.

DENTAN, E. Etude microscopique du développement et de la maturation du grain de café. In: INTERNATIONAL SCIENTIFIC COLLOQUIUM ON COFFEE, 11., 1985, Lomé. **Proceedings...** Paris: International Scientific Association on Coffee, 1985. p. 381-198.

DUNWELL, J. M. Transgenic approaches to crop improvement. **Journal of Experimental Botany**, Oxford, v. 51, n. 1, p. 487-496, Feb. 2000.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Cenargen**. Disponível em: <<https://alanine.cenargen.embrapa.br/cafEST/>>. Acesso em: 10 dez. 2012a.

_____. **Genoma café**. Disponível em: <<http://www.lge.ibi.unicamp.br/cafe/>>. Acesso em: 10 dez. 2012b.

FAZUOLI, L. C. et al. Seleção antecipada e sua eficiência no café Icatu. In: SIMPÓSIO BRASILEIRO DE PESQUISAS DOS CAFÉS DO BRASIL, 1., 2000, Poços de Caldas. **Anais...** Poços de Caldas: Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café, 2000. p. 576-582.

FERNANDEZ, D. et al. 454-pyrosequencing of *Coffea Arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. **Molecular Plant Pathology**, London, v. 13, n. 1, p. 17-37, Jan. 2012.

FLEXAS, J.; MEDRANO, H. Drought-inhibition of photosynthesis in C₃ plants: stomatal and non-stomatal limitations revisited. **Annals of Botany**, London, v. 89, n. 2, p. 183-189, Feb. 2002.

GEROMEL, C. et al. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. **Journal of Experimental Botany**, Oxford, v. 57, n. 12, p. 3243-3258, Sept. 2006.

GONÇALVES, W. Melhoria do cafeeiro visando resistência a nematoide. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3., 1999, Lavras. **Anais...** Lavras: UFLA, 1999. 1 CD-ROM.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current Opinion in Plant Biology**, Oxford, v. 5, n. 2, p. 122-127, Apr. 2002.

GUZZO, S. D.; HARAKAVA, R.; TSAI, S. M. Identification of coffee genes expressed during systemic acquired resistance and incompatible interaction with *Hemileia vastatrix*. **Journal of Phytopathology**, Berlin, v. 157, n. 10, p. 625-638, Oct. 2009.

HERRERA, J. C. et al. Introgression into the allotetraploid coffee (*Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* x *C. canephora*). **Theoretical and Applied Genetics**, Berlin, v. 104, n. 4, p. 661-668, Mar. 2002.

HÖFTE, H. et al. An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. **Plant Journal**, Oxford, v. 5, n. 4, p. 611-611, Apr. 1994.

HORAN, K. et al. Annotating genes of known and unknown function by large-scale coexpression analysis. **Plant Physiology**, Bethesda, v. 147, n. 1, p. 47-51, May 2008.

HOULLOU-KIDO, L. M. et al. Bioinformática aplicada ao estudo de genomas expressos (EST) em plantas. In: FIGUEIREDO, M. do V. B. et al. (Ed.). **Bioteecnologia aplicada à agricultura: textos de apoio e protocolos experimentais**. Brasília: EMBRAPA Informação Tecnológica; Recife: Instituto Agrônômico de Pernambuco, 2010. p. 257-332.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. **The physical science basis: summary for policymakers**. Geneva, 2007. Disponível em: <<http://www.ipcc.ch>>. Acesso em: 23 jan. 2013.

INTERNATIONAL COFFEE ORGANIZATION. **Trade statistics**. Disponível em: <http://www.ico.org/trade_statistics.asp?section=Statistics>. Acesso em: 23 jan. 2013.

JOËT, T. et al. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. **New Phytologist**, Cambridge, v. 182, n. 1, p. 146-162, Apr. 2009.

KNIGHT, H.; KNIGHT, M. R. Abiotic stress signalling pathways: specificity and cross-talk. **Trends in Plant Science**, Oxford, v. 6, n. 6, p. 262-267, June 2001.

KOONIN, E. V. Orthologs, paralogs, and evolutionary genomics. **Annual Review of Genetics**, Palo Alto, v. 39, p. 309-338, Dec. 2005.

LASHERMES, P.; BERTRAND, B.; ETIENNE, H. Breeding coffee (*Coffea arabica*) for sustainable production. In: JAIN, S. M.; PRIYADARSHAN, P. M. (Ed.). **Breeding plantation tree crops**. New York: Springer, 2009. p. 525-544.

LASHERMES, P. et al. Analysis of alien introgression in coffee tree (*Coffea arabica* L.). **Molecular Breeding**, Dordrecht, v. 27, n. 2, p. 223-232, Feb. 2011.

_____. Molecular characterisation and origin of the *Coffea arabica* L. genome. **Molecular and General Genetics**, New York, v. 261, n. 2, p. 259-266, Mar. 1999.

LEROY, T. et al. Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. **Theoretical and Applied Genetics**, Berlin, v. 111, n. 6, p. 1032-1041, Oct. 2005.

_____. Genetics of coffee quality. **Brazilian Journal of Plant Physiology**, Piracicaba, v. 18, n. 1, p. 229-242, 2006.

LIMA, A. L. S. et al. Photochemical responses and oxidative stress in two clones of *Coffea canephora* under water deficit conditions. **Environmental and Experimental Botany**, Oxford, v. 47, n. 3, p. 239-247, May 2002.

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**, Berlin, v. 112, n. 1, p. 114-130, Dec. 2005.

LOON, L. C. van; REP, M.; PIETERSE, C. M. Significance of inducible defense-related proteins in infected plants. **Annual Review of Phytopathology**, Palo Alto, v. 44, p. 135-162, Sept. 2006.

MA, J. B. et al. Identification of expressed genes during compatible interaction between stripe rust (*Puccinia striiformis*) and wheat using a cDNA library. **BMC Genomics**, London, v. 10, p. 586-590, Dec. 2009.

MARRACCINI, P. et al. Differentially expressed genes and proteins upon drought stress in tolerant and sensitive genotypes of *Coffea canephora*. **Journal of Experimental Botany**, Oxford, v. 63, n. 11, p. 4191-4212, June 2012.

_____. *RBCS1* expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. **BMC Plant Biology**, London, v. 11, n. 85, p. 1-23, May 2011.

MAURIN, O. et al. Towards a phylogeny for *Coffea* (*Rubiaceae*): identifying well-supported lineages based on nuclear and plastid DNA sequences. **Annals of Botany**, London, v. 100, n. 7, p. 1565-1583, Dec. 2007.

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**, London, v. 11, p. 30-34, Feb. 2011.

_____. Isolation and characterization of *Coffea* genes induced during coffee leaf miner (*Leucoptera coffeella*) infestation. **Plant Science**, Shannon, v. 169, n. 2, p. 351-360, Apr. 2005.

MONTAGNON, C.; CUBRY, P.; LEROY, T. Amélioration génétique du caféier *Coffea canephora* Pierre: connaissances acquises, stratégies et perspectives. **Cahiers Agricultures**, Madrid, v. 21, n. 2/3, p. 143-153, 2012.

MONTAGNON, C.; LEROY, T. Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de Côte-d'Ivoire appartenant à différents groupes génétiques. **Café Cacao Thé**, Paris, v. 37, n. 3, p. 179-190, 1993.

MONTOYA, G. et al. Sequence analysis from leaves, flowers and fruits of *Coffea arabica* var. Caturra. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 21., 2006, Montpellier. **Proceedings...** Montpellier: ASIC, 2006. 1 CD-ROM.

MUELLER, L. A. et al. The tomato sequencing project, the first cornerstone of the international *Solanaceae* Project (SOL). **Comparative and Functional Genomics**, Chichester, v. 6, n. 3, p. 153-158, Apr. 2005.

NAGARAJ, S. H.; GASSER, R. B.; RANGANATHAN, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. **Briefings in Bioinformatics**, London, v. 8, n. 1, p. 6-21, Jan. 2007.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **dbEST**. Disponível em: <<http://www.ncbi.nlm.nih.gov/dbEST/>>. Acesso em: 10 dez. 2012.

NELSON, G. C. et al. **Climate change**: impact on agriculture and costs of adaptation. Washington: International Food Policy Research Institute, 2009. 30 p.

PEZZOPANE, J. R. M. et al. Escala para avaliação de estádios fenológicos do cafeeiro arábica. **Bragantia**, Campinas, v. 62, n. 3, p. 499-505, 2003.

PINHEIRO, H. A. et al. Drought tolerance is associated with rooting depth and stomatal control of water use in clones of *Coffea canephora*. **Annals of Botany**, London, v. 91, n. 1, p. 101-108, Mar. 2005.

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**, Berlin, v. 276, n. 5, p. 436-449, Nov. 2006.

PRIVAT, I. et al. Differential regulation of grain sucrose accumulation and metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) revealed through gene expression and enzyme activity analysis. **New Phytologist**, Cambridge, v. 178, n. 4, p. 781-797, Mar. 2008.

RIBAS, A. F.; PEREIRA, L. F. P.; VIEIRA, L. G. Genetic transformation of coffee. **Brazilian Journal of Plant Physiology**, Piracicaba, v. 18, n. 1, p. 83-94, Mar. 2006.

SALES, R. M. O. B.; BARBOSA, A. V.; SILVA, F. R. O banco de dados do "Projeto Genoma Café" na EMBRAPA Recursos Genéticos e Biotecnologia (CENARGEN). In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 5., 2007, Águas de Lindóia. **Anais...** Águas de Lindóia: SBC, 2007. 1 CD-ROM.

SALMONA, J. et al. Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. **Plant Molecular Biology**, Berlin, v. 66, n. 1/2, p. 105-124, Jan. 2008.

SEKI, M. et al. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. **Plant Cell**, Rockville, v. 13, n. 1, p. 61-72, Jan. 2001.

STERKY, F.; LUNDEBERG, J. Sequence analysis of genes and genomes. **Journal of Biotechnology**, Amsterdam, v. 76, n. 1, p. 1-31, Jan. 2000.

TAUTZ, D.; DOMAZET-LOSO, T. The evolutionary origin of orphan genes. **Nature Reviews Genetics**, London, v. 12, n. 10, p. 692-702, Oct. 2011.

VIDAL, R. O. et al. A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. **Plant Physiology**, Bethesda, v. 154, n. 3, p. 1053-1066, Nov. 2010.

VIEIRA, L. G. E. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, Piracicaba, v. 18, n. 1, p. 95-108, Mar. 2006.

VOSSSEN, H. A. M. van der. Coffee breeding practices. In: CLARKE, R. J.; VITZTHUM, O. G. (Ed.). **Coffee: recent developments, agronomy**. London: Blackwell Science, 2001. v. 1, p. 184-201.

WANG, W. X.; VINOCCUR, B.; ALTMAN, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. **Planta**, Berlin, v. 218, n. 1, p. 1-14, Nov. 2003.

YAMAMOTO, K.; SASAKI, T. Large-scale EST sequencing in rice. **Plant Molecular Biology**, Berlin, v. 35, n. 1/2, p. 135-144, Sept. 1997.

ZHANG, W. Y. et al. A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. **PLOS One**, San Francisco, v. 6, n. 3, p. 1-7, Mar. 2011.

CAPÍTULO 2

IDENTIFICAÇÃO E CARACTERIZAÇÃO DA EXPRESSÃO *IN SILICO* DOS *NO HITS* PRESENTES NAS BASES DE DADOS DE CAFÉ

1 INTRODUÇÃO

Um dos requisitos primordiais para o uso das novas ferramentas biotecnológicas é a identificação e caracterização funcional de genes. Atualmente uma parcela bastante representativa dos projetos genoma desenvolvida é utilizada para identificação de genes, identificação de transcritos alternativos e caracterização de proteomas (HORAN et al., 2008). Técnicas para se analisar a estrutura e função de genes, em larga escala, proporcionam uma enorme produção de informações e a geração de bancos de dados de sequências de DNA e proteínas.

As sequências vêm sendo depositadas, em sua maioria, em bancos de dados públicos, facilitando o acesso a essas informações por quaisquer grupos de pesquisa no mundo inteiro. Esses bancos de dados genômicos contêm tanto as sequências de DNA como sequências de etiquetas expressas (EST) (NAGARAJ; GASSER; RANGANATHAN, 2007). Contudo, a comparação de uma sequência desconhecida com um vasto número de ESTs tem se tornado um instrumento importante e poderoso na pesquisa genética como foi relatado no Capítulo 1.

Os bancos de dados de café estão sendo ampliados a cada dia com os estudos dos projetos genoma, inicialmente com o Projeto Brasileiro do Genoma Café – PBGC (VIEIRA et al., 2006). Todavia, com o advento das novas tecnologias de sequenciamento (ZHANG et al., 2011) há uma geração enorme de dados que necessitam ser estudados e analisados. Uma parcela bastante significativa destes dados são referentes às sequências desconhecidas, os *no hits*.

Estas sequências são tratadas como órfãs por Tautz e Domazet-Lošo (2011), e embora os *no hits* tenham sido pouco estudados em grande detalhe funcional, até agora, está se tornando cada vez mais evidente que eles têm contribuído substancialmente para a evolução adaptativa de organismos

biológicos. Os genes órfãos surgem a todo o momento por processos evolutivos como, por exemplo, mutações e produção de genes *de novo* (XIE et al., 2012).

Tais processos podem gerar novas sequências gênicas capazes de serem transcritas e talvez traduzidas, percorrendo a etapa evolutiva de seleção natural. As possíveis proteínas geradas no processo evolutivo podem estar implicadas na adaptação aos estresses bióticos e abióticos das plantas.

As informações fornecidas com o estudo dos transcriptomas são de grande importância não apenas para a identificação de novos genes, mas também por fornecer informações que indiquem sua provável função biológica, devido a sua condição de expressão, assim como as rotas metabólicas associadas. As condições fisiológicas podem inclusive induzir um *splicing* alternativo, conferindo a um gene específico uma grande variação funcional, de acordo com a montagem dos exons.

Sendo assim, o objetivo do trabalho neste capítulo foi identificar e caracterizar funcionalmente os genes órfãos (*no hits*) nas espécies *Coffea arabica* e *Coffea canephora*, por meio de uma análise exploratória da expressão gênica *in silico* em diferentes tecidos e condições de estresse (biótico e abiótico).

2 MATERIAL E MÉTODOS

2.1 Identificação dos *NO HITS*

2.1.1 Busca nas bases de dados e alinhamentos

Como resultado do Consórcio Internacional de Sequenciamento do Genoma de *C. canephora*, foram preditas 25.574 sequências protéicas (WINCKER et al., 2011). Estas sequências foram utilizadas neste trabalho como referência para a busca por homologia nas bases de dados do NCBI. A validação dos dados foi realizada pela pesquisa por meio da ferramenta BLASTp. As sequências foram contrastadas com o banco de dados do NCBI NR no GenBank (BASIC LOCAL ALIGNMENT SEARCH TOOL - BLAST, 2012) utilizando o software Geneious (KEARSE et al., 2012). Os alinhamentos resultantes foram filtrados por meio de um valor de E-value de $1e^{-1}$ como limite para identidade. As sequências que não possuíram nenhuma similaridade com aquelas conhecidas nos bancos de dados do GenBank (BLAST, 2012) foram denominadas “*no hits*”, estas sequências foram então extraídas e utilizadas como alvo para as análises de expressão *in silico*.

2.2 Composição das Bases de dados das bibliotecas para as análises *in silico*

Os dados obtidos para este estudo são provenientes de dois tipos de sequenciamento:

Primeiramente, foram utilizados os dados gerados com a tecnologia de sequenciamento do tipo Sanger, que utiliza o método enzimático com base na síntese de DNA *in vitro*, na presença de nucleotídeos trifosfatados terminadores de cadeia (SANGER; NICKLEN; COULSON, 1977). As sequências utilizando

este tipo de sequenciamento são provenientes das bibliotecas relatadas em Lin et al. (2005), Poncet et al. (2006) e Vieira et al. (2006).

Segundo, foram utilizados os 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 utilizando esta tecnologia foram geradas a partir de extratos de RNAs de raízes de *Coffea 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., 2011). Foram utilizados também extratos de RNAs de meristemas de *C. arabica* das cultivares Rubi (considerado sensível à seca) nas condições I e NI e IAPAR59 (considerado tolerante à seca) nas condições I e NI (VIDAL et al., 2012) (Ver Capítulo 3 para as especificações do experimento). O sequenciador 454 baseia-se na técnica de detecção do pirofosfato (PPi) durante a síntese de DNA pela DNA polimerase (ZHANG et al., 2011).

2.3 Bibliotecas geradas com o sequenciamento

As análises de expressão *in silico* foram realizadas separadamente de acordo com as tecnologias de sequenciamento utilizadas para a geração dos dados das bibliotecas. Como o número de informações é muito maior com o sequenciamento utilizando a tecnologia 454 quando comparados com o sequenciamento com Sanger, os resultados poderiam ser superestimados se os dados fossem agrupados.

2.3.1 Bibliotecas geradas com o sequenciamento do tipo Sanger

Foram utilizados dados das bibliotecas provenientes das bases de dados do café, onde cerca de 10.420 sequências utilizadas neste trabalho são

proveniente de 2 bibliotecas de cDNA, de folha e fruto de *C. canephora* em diferentes estádios de maturação (PONCET et al., 2006), 32 bibliotecas de *C. arabica* e 3 bibliotecas de *C. canephora*, cerca de 163.623 sequências utilizadas foram provenientes do PBGC (VIEIRA et al., 2006), 1 biblioteca de folha e 4 bibliotecas de frutos em diferentes estádios e tecidos de *C. canephora* também foram utilizadas neste trabalho, cerca de 30.305 sequências foram provenientes destas bibliotecas (LIN et al., 2005). Outras 16 bibliotecas (24.115 sequências) provenientes de flores, folhas, frutos e raiz de *C. arabica* buscadas nas bases de dados do café em dbNCBI (dados não publicados) complementaram os dados das bibliotecas de café utilizadas com a tecnologia Sanger, gerando um total de 228463 reads.

Os dados de sequências destas bibliotecas utilizadas foram reunidos coerentemente em conjuntos contendo 13 bibliotecas mistas, agrupadas de acordo com a espécie (sendo CC bibliotecas provenientes da espécie *C. canephora* e CA bibliotecas provenientes de *C. arabica*) e tecidos (onde bibliotecas de tecidos iguais foram agrupadas em uma biblioteca maior). As novas bibliotecas foram nomeadas: CC-Folha, CC- Callus, CC- Fruto, CA-Folhas, CA- Raiz, CA- Embriões, CA- Fruto, CA- Callus, CA- Susp. Cel., CA-Plântulas, CA-Flores, CA- Ramos e CA-Pool tecidos e utilizadas nas análises de expressão dos *no hits in silico*. A frequência dos reads de cada biblioteca foi calcula e anotada. (Tabela 1)

Tabela 1 Bibliotecas utilizadas nas análises de expressão *in silico*. Os dados gerados utilizando a tecnologia Sanger.

Nome da biblioteca	Tecidos/descrição	Número de reads	Referência
CC- Folha	Folhas jovens e adultas de <i>C. canephora</i>	18.155	Vieira et al. (2006) Poncet et al. (2006) Lin et al. (2005)
CC- Callus	Calos embriogênicos de <i>C. canephora</i>	6.888	Vieira et al. (2006)
CC- Fruto	Frutos em diferentes estádios de <i>C. canephora</i>	28.537	Poncet et al. (2006) Lin et al. (2005)
CA- Folha	Folhas jovens e adultas + folhas infectadas com bichomineiro e ferrugem (<i>C. arabica</i>)	29.641	Vieira et al. (2006)
CA- Raiz	Raiz + Raiz com nematoide + Raiz com acibenzolar-S-metil (<i>C. arabica</i>)	3.033	Vieira et al. (2006)
CA- Embriões	Sementes em germinação (sementes inteiras e embriões zigóticos)	7.905	Vieira et al. (2006)
CA- Fruto	Frutos em diferentes estádios (<i>C. arabica</i>)	20.555	Vieira et al. (2006)
CA- Callus	Calo embriogênico e calo não embriogênico com e sem 2,4D de <i>C. arabica</i>	18.726	Vieira et al. (2006)
CA- Susp. Cel.	Suspensão de células tratadas com acibenzolar-S-metil, ou NaCl ou brassinosteroides (<i>C. arabica</i>)	27.638	Vieira et al. (2006)
CA- Plântulas	Plântulas e folhas tratadas com ácido araquidônico + Hipocótilos tratados com acibenzolar-S-metil (<i>C. arabica</i>)	14.360	Vieira et al. (2006)
CA- Flores	Flores em diferentes estádios de desenvolvimento (<i>C. arabica</i>)	22.944	Vieira et al. (2006)
CA- Ramos	Ramos infectados com <i>Xylella spp.</i> (<i>C. arabica</i>)	8.292	Vieira et al. (2006)
CA- Mix Pool	Pool de tecidos de <i>C. arabica</i>	21.789	Vieira et al. (2006)

1º: Foi realizado a análise da expressão *in silico* dos *no hits* primeiramente com relação às bibliotecas de *C. arabica* utilizando a tecnologia de sequenciamento do tipo Sanger, estas bibliotecas foram: CA- Folha, CA- Raiz, CA- Embriões CA- Fruto, CA- Callus, CA- Susp. Cel., CA- Plântulas, CA-Flores, CA- Ramos e CA-Pool tecidos.

2º: Após, foi realizada uma análise de expressão *in silico* dos *no hits* com relação às bibliotecas de *C. canephora* utilizando a tecnologia de sequenciamento do tipo Sanger, estas bibliotecas foram: CC-Folha, CC- Callus, CC- Fruto.

Foi possível separar e analisar mais a fundo os dados das sequências das bibliotecas de CC-Folha e CC-Fruto. A biblioteca CC-Folha foi subdividida em CC-Folha controle, que contém dados de bibliotecas de folhas não estressadas hidricamente (PONCET et al., 2006) e CC- Folha estresse, que contém dados de bibliotecas de folhas estressadas hidricamente (VIEIRA et al., 2006). (Tabela 3)

3º: Foi realizada então uma análise de expressão *in silico* dos *no hits* com relação às bibliotecas de *C. canephora* de CC-Folha controle e CC- Folha estresse, utilizando a tecnologia de sequenciamento do tipo Sanger.

No caso da biblioteca de CC-Fruto foi possível comparar os dados de frutos de *C. canephora* em diferentes estádios de desenvolvimento. No estágio inicial foram agrupados ESTs provenientes de frutos inteiros coletados no período de 18 (126 dias) a 22 (154 dias) semanas após a floração. No estágio intermediário ESTs de endosperma e perisperma de frutos coletados 30 semanas (210 dias) após a floração foram agrupados. No estágio tardio, ESTs de endosperma e perisperma de frutos coletados 42 (294 dias) e 46 (322) semanas após a floração foram agrupados. Os ESTs do pericarpo dos frutos nos diferentes estádios também foram agrupados (LIN et al., 2005). A biblioteca CC- Fruto foi então subdividida em CC-Fruto est. inicial, CC-Frutos est. entrem. e CC-Frutos est. tardio (Tabela 2)

Tabela 2 Subdivisões das bibliotecas CC-Folhas e CC- Fruto.

Nome da biblioteca	Subdivisão	Tecidos/descrição	Número de reads	Referência
CC- Folha	CC- Folha controle	Folhas jovens e adultas de <i>C. canephora</i>	9.188	Poncet et al. (2006)
	CC- Folha estresse	Folhas de <i>C. canephora</i> estressadas hidricamente	5.967	Vieira et al. (2006)
CC- Fruto	CC-Frutos est. inicial	Frutos no estágio inicial de desenvolvimento	9.843	Lin et al. (2005)
	CC-Frutos est. interm.	Frutos no estágio intermediário de desenvolvimento	10.077	
	CC-Frutos est. tardio	Frutos no estágio final de desenvolvimento	9.096	
	CC-Frutos Pericarpo	Pericarpo dos frutos e vários estádios de desenvolvimento	8.956	

4º: Após, foi realizada uma análise de expressão *in silico* dos *no hits* com relação às bibliotecas de frutos de *C. canephora* nos diferentes estádios de desenvolvimento utilizando a tecnologia de sequenciamento do tipo Sanger, estas bibliotecas foram: CC-Fruto est. inicial, CC-Frutos est. interm. e CC-Frutos est. tardio.

2.3.2 Bibliotecas geradas com o sequenciamento do tipo 454

Os dados das sequências geradas com a tecnologia do tipo 454 foram agrupados em 12 bibliotecas, sendo 8 delas referentes a 3.858.526 sequências de raízes de *C. canephora* provenientes dos clones 14, 22, 73 e 120 nas condições irrigado (I) e não-irrigado (NI) e 4 bibliotecas referentes a 993.260 sequências de meristemas de *C. arabica* referentes aos cultivares Rubi e IAPAR59 (I59) também nas condições I e NI (Tabela 3).

Tabela 3 Bibliotecas utilizadas nas análises de expressão *in silico*. Os dados gerados utilizando a tecnologia 454.

Tecidos	Nome da biblioteca	Descrição	Número de reads	Referência
Raiz	CC-14I	Raiz de <i>C.canephora</i> irrigado	559.079	Dados não publicados
	CC-14NI	Raiz de <i>C.canephora</i> não-irrigado	603.473	
	CC-22I	Raiz de <i>C.canephora</i> irrigado	560.932	
	CC-22NI	Raiz de <i>C.canephora</i> não-irrigado	630.779	
	CC-73I	Raiz de <i>C.canephora</i> irrigado	499.501	
	CC-73NI	Raiz de <i>C.canephora</i> não-irrigado	539.102	
	CC-120I	Raiz de <i>C.canephora</i> irrigado	450.456	
	CC-120NI	Raiz de <i>C.canephora</i> não-irrigado	554.306	
Meristema	CA-RUBI I	Meristema de <i>C.arabica</i> irrigado	345.765	Dados não publicados
	CA-RUBI NI	Meristema de <i>C.arabica</i> não-irrigado	230.066	
	CA-I59 I	Meristema de <i>C.arabica</i> irrigado	282.167	
	CA-I59 NI	Meristema de <i>C.arabica</i> não-irrigado	135.262	

5°: Foi realizado a análise da expressão *in silico* dos *no hits* com relação às bibliotecas de raízes de *C. canephora* utilizando a tecnologia de sequenciamento do tipo 454, estas bibliotecas foram: CC-14 I, CC-14 NI, CC-22 I, CC-22 NI, CC-73 I, CC-73 NI, CC-120 I e CC-120 NI.

6º: Foi realizado a análise da expressão *in silico* dos *no hits* com relação às bibliotecas de meristema de *C. arabica* utilizando a tecnologia de sequenciamento do tipo 454, estas bibliotecas foram: CA-Rubi I, CA-Rubi NI, CA-I59 I e CA-I59 NI.

2.4 Northern Eletrônico

Para as análises de expressão *in silico* (*Northern* eletrônico) foi utilizado o pacote DNASTar, software Qseq, versão 7.0. Foi realizado então um mapeamento dos *no hits*, encontrados a partir da busca por homologia, em relação aos *reads* das bibliotecas, gerando assim os dados de expressão dos *no hits* em cada biblioteca. Os dados foram previamente normalizados com a estandardização da expressão dos *no hits* sempre relativos à expressão do gene da ubiquitina (*CcUBQ10*), usado como um calibrador interno de referência, este mesmo gene foi utilizado como controle endógeno para as análises de qPCR apresentadas no Capítulo 3. A expressão do gene da ubiquitina é apresentada uniformemente entre as bibliotecas e é mostrada ao final de todos os gráficos, juntamente com a expressão dos genes *no hits*. As sequências foram reorganizadas de acordo com o seu grupo de classificação, e os seus dados foram lançados em uma matriz relativa dos genes *versus* bibliotecas. As sequências dos *no hits* e as bibliotecas foram agrupadas em uma estrutura hierárquica e visualizados pelo programa TreeView (EISEN et al., 1998). Os resultados de expressão foram apresentados em um espectro de escala de cores. Quando não houve expressão ou foi negativo a região foi corada com o tom pastel, sendo gradualmente aumentado até atingir o vermelho intenso, que representa o grau máximo positivo de expressão.

3 RESULTADOS E DISCUSSÃO

A busca por homologia na pesquisa por BLASTp das 25.574 sequências protéicas contra o banco de dados do GenBank geraram 24.006 sequências homólogas (*hits*) e 1.568 sequências que não possuíam nenhuma similaridade (filtrados com um valor de E-value de $\leq 1e^{-1}$ como limite para identidade) com aquelas conhecidas no banco de dados do GenBank (*no hits*), estes os quais foram utilizados para as análises posteriores. Estes *no hits* estão distribuídos por todos os cromossomos de café possuindo um número representante de dados desconhecidos e que precisam não só ser identificados e catalogados, mas também atribuir à função destes genes e produtos gênicos, compreendendo as interligações entre DNA, RNA e proteínas (Figura 1).

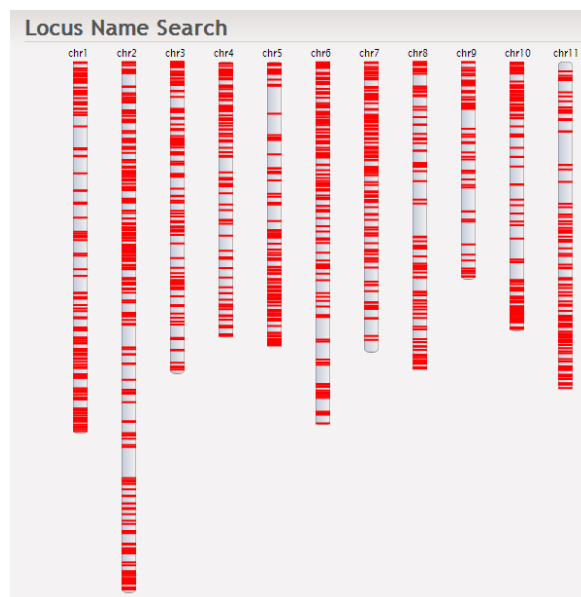


Figura 1 Representação esquemática de cada locus dos *no hits* encontrados, distribuídos nos 11 cromossomos de café.

3.1 Análise de expressão gênica *in silico* nas bibliotecas de *C. arabica* (CA) utilizando a tecnologia Sanger

Das 1.568 sequências de *no hits* utilizadas para as análises *in silico* nas bibliotecas de *C. arabica* (CA) utilizando a tecnologia de sequenciamento do tipo Sanger, 116 delas apresentaram perfis de expressão nestas bibliotecas, porém 90 *no hits* foram selecionado para apresentação na Figura 2. É possível notar uma expressão diferencial dos *no hits* com relação às bibliotecas apresentadas. Parece haver uma expressão mais intensa dos *no hits* na biblioteca de CA-Plântulas. Sendo que o *no hit* mais expresso foi o GSCOCT0003359000 seguido do GSCOCT00033656001, na biblioteca de CA-Plântulas, o *no hit* GSCOCT000042747001 é o 3º mais expresso, presente com mais intensidade da biblioteca CA-Suspensão celular, em 4º lugar segue o *no hit* GSCOCT00005517001, sendo mais expresso em CA-Raiz, após, segue o GSCOCT00038657001 e o GSCOCT00030926001, respectivamente.

O *no hit* GSCOCT00030926001 foi expresso em todos os tecidos. No entanto alguns genes foram específicos para um determinado tecido ou apresentou uma expressão muito superior com relação aos demais, como no caso do GSCOCT00015847001, que parece ser específico e altamente expresso em CA-Plântulas, O *no hit* GSCOCT00022858001 foi específico da biblioteca de CA-Fruto. Enquanto que o intervalo dos genes GSCOCT00014522001 até o GSCOCT00039348001 presentes na parte média inferior da Figura 2 foram específicos da biblioteca de CA-Callus. O *no hit* GSCOCT00007128001 parece ser expresso pelo menos 10x mais em CA-Fruto do que nos demais tecidos. Em CA-Raiz destaca-se o *no hit* GSCOCT00005517001 e o GSCOCT00033656001 presentes também em CA-Plântulas com intensidade. A biblioteca de CA-Ramo apresentou alta expressão para o *no hit* GSCOCT00030926001, em CA-Embrião o *no hit* mais expresso foi o GSCOCT00036640001, em CA-Flor foi o

GSCOCT00005517001, no entanto com pouca intensidade e em CA-Folha o GSCOCT00033656001 foi o mais expresso.

A biblioteca CA-Plântulas é composta de plântulas e folhas tratadas com ácido araquidônico e hipocótilos tratados com acibenzolar-S-metil de *C. arabica*. O acibenzolar-S-metil é um fungicida do grupo químico dos benzothiadiazoles, que tem sido testado como um indutor de resistência em plantas. Não age diretamente sobre o agente patogênico, mas ativa naturalmente as defesas da planta, deixando a mesma menos suscetível a doenças e a pragas.

O ácido araquidônico nas plantas servem como substrato para a atividade das lipoxigenases envolvidas na rota do mecanismo de defesa das plantas às lesões ou a patógenos. Assim, estes genes podem estar implicados na resposta da planta à patógenos ou às lesões, participando ativamente das rotas metabólicas que levam a estas respostas.

É interessante notar que outras bibliotecas também possuem o tratamento com o acibenzolar-S-metil, como no caso de CA-Raiz que também contém ESTs de raízes tratadas com o acibenzolar-S-metil e CA-Suspensão celular que possuem ESTs de Suspensão de células tratadas com acibenzolar-S-metil, ou NaCl ou brassinosteroides. Estas três bibliotecas possuem genes altamente expressos concomitantemente, como no caso do gene GSCOCT00033590001, GSCOCT00033656001 e GSCOCT00030926001, podendo atuar na defesa das plantas em diversos tecidos.

Os *no hits* com potencial participação no estresse biótico selecionados para as análises *in vivo* (qPCR) no capítulo 3 foram GSCOCT00033590001, GSCOCT00033656001 e GSCOCT00030926001, estes estão marcados com asterisco de cor vermelha na Figura 2.

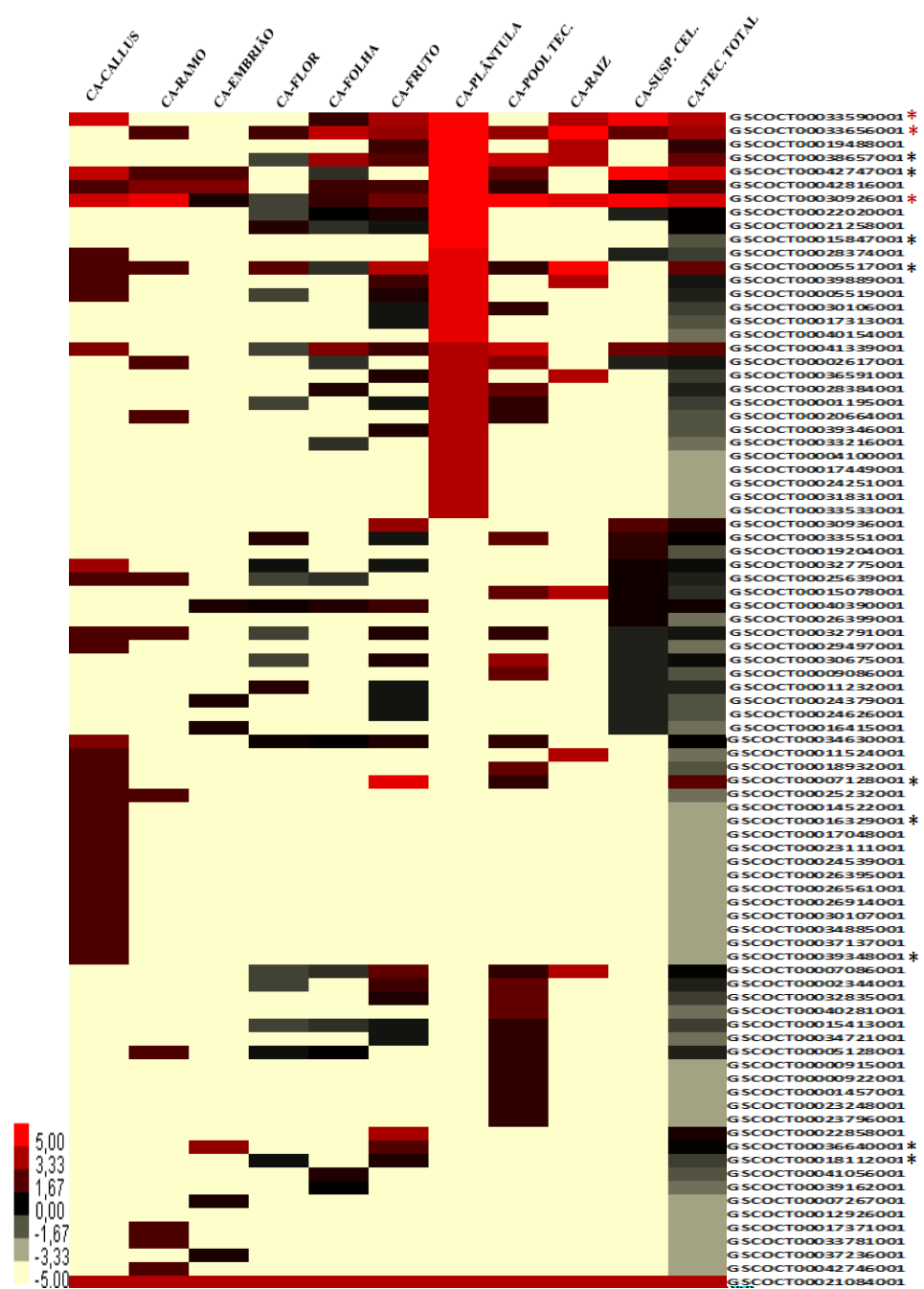


Figura 2 Expressão gênica *in silico* dos *no hits* nas bibliotecas de *C. arabica* (CA) utilizando a tecnologia Sanger. (*) *no hits* citados no texto. (*) *no hits* citados no texto e selecionados para qPCR.

3.2 Análise de expressão gênica *in silico* das bibliotecas de *C. canephora* (CC) utilizando a tecnologia Sanger

Das 1.568 sequências de *no hits* gerados na busca por similaridade, 187 delas apresentaram perfis de expressão nas bibliotecas de CC-Callus, CC-Frutos e CC-Folhas utilizando a tecnologia de sequenciamento do tipo Sanger, 81 *no hits* estão apresentados na Figura 3 e estes parecem ter uma expressão diferencial entre os tecidos de *C. canephora* utilizados. O *no hit* mais expresso entre os tecidos estudados foi o GSCOCT00026561001 seguido do GSCOCT00036640001, GSCOCT00042747001, GSCOCT00019488001, GSCOCT00024379001, respectivamente.

Em CC-Callus o *no hit* mais expresso foi o GSCOCT00026561001 seguido dos GSCOCT00024379001, GSCOCT00022020001, GSCOCT00022223001, GSCOCT00030107001 e GSCOCT00037137001 que possuíram a mesma intensidade.

Alguns *no hits* parecem ser específicos para a biblioteca de CC-Callus, destacando o GSCOCT00026561001, GSCOCT00022223001, GSCOCT00030107001 e GSCOCT00037137001. Comparando os perfis de expressão específicos para callus nas análises de *C. arabica* (Figura 2) e *C. canephora* (Figura 3), os *no hits* GSCOCT00026914001 e GSCOCT00030107001 estão presentes especificamente na biblioteca de callus, tanto nas análises de CA (Figura 2) quanto nas análises de CC (Figura 3). Estes *no hits* parecem ser genes específicos para este tecido tanto nas espécies de *C. arabica*, quanto nas espécies de *C. canephora*, isso indica que *estes no hits* parecem pertencer ao subgenoma CaCc introgrado em *C. arabica* como relatado em Vidal et al. (2010).

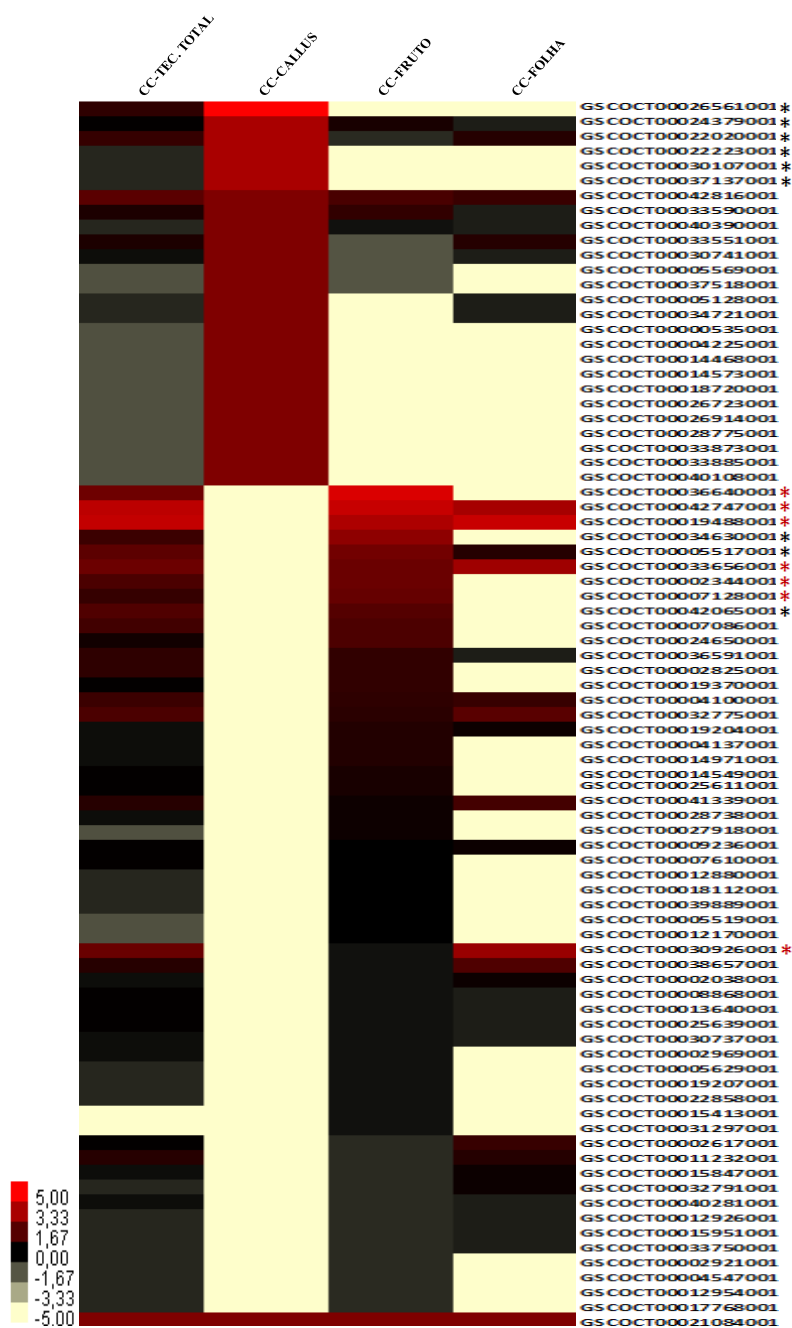


Figura 3 Expressão gênica *in silico* dos *no hits* nas bibliotecas de *C. canephora* (CC) utilizando a tecnologia Sanger. (*) *no hits* citados no texto. (*) *no hits* citados no texto e selecionados para qPCR.

Em CC-Folha, o *no hit* mais expresso foi o GSCOCT00019488001 seguido do GSCOCT00042747001, do GSCOCT00033656001 e do GSCOCT00030926001, respectivamente. Em folhas não houve especificidade dos genes neste tecido, no entanto alguns genes apresentaram expressão maior em CC-Folhas quando comparados aos demais tecidos.

Já na biblioteca de CC-Fruto (Figura 3) o *no hit* mais expresso foi o GSCOCT00036640001 seguido do GSCOCT00042747001, GSCOCT00019488001, GSCOCT00034630001, GSCOCT00005517001 e GSCOCT00033656001, respectivamente. Alguns *no hits* parecem ser específicos deste tecido, destacando o GSCOCT00036640001, GSCOCT00034630001, GSCOCT00002344001, GSCOCT00007128001 e GSCOCT00042065001.

Quando comparamos as bibliotecas de folhas estressadas e não-estressadas hidricamente presentes na Figura 4, os *no hits* mais expressos em CC-Folha se destacam com relação à condição ou não de estresse de seca. É possível ressaltar *no hits* que parecem ser altamente induzidos com o estresse de seca como no caso do GSCOCT00033656001 e GSCOCT00030926001.

É possível perceber também na Figura 4, *no hits* que parecem ser altamente reprimidos na condição de estresse de seca, esta situação é ressaltada para os GSCOCT00019488001 e GSCOCT00042747001, seguidos de GSCOCT00032775001 e GSCOCT00041339001, respectivamente. Estes *no hits* já apareceram com uma acentuada expressão nas bibliotecas de CA-Plântulas (visualizado na Figura 2), com exceção do GSCOCT00032775001. Podendo relacioná-los a diversos estresses simultaneamente. Além do que é possível inferir que os *no hits* GSCOCT00019488001, GSCOCT00042747001 e GSCOCT00041339001, dentre outros parecem pertencer ao subgenoma de *C. canephora* (*CaCc*) em *C. arabica* como descrito em Vidal et al. (2010).

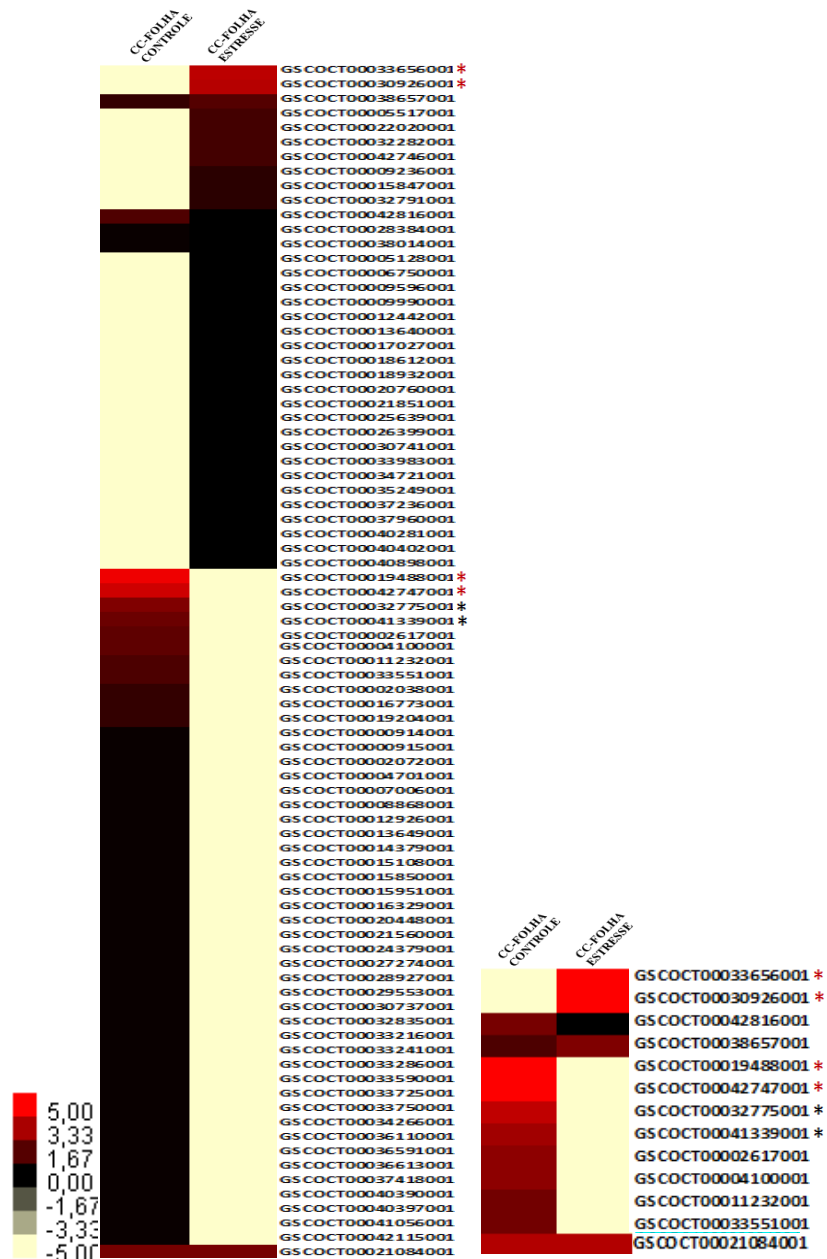


Figura 4 Expressão gênica *in silico* dos *no hits* nas bibliotecas de folhas de *C. canephora* (CC) com e sem estresse de seca, utilizando a tecnologia Sanger. À direita, um enfoque nos genes mais expressos. (*) genes citados no texto. (*) genes citados no texto e selecionados para qPCR.

O resultado do *Northern eletrônico* para os frutos de *C. canephora* em diferentes estádios de desenvolvimento é apresentado na Figura 5. Os genes presentes em CC-Fruto (Figura 5) apresentam indícios de potencial participação do processo de maturação dos frutos de *C. canephora*.

É possível observar na Figura 5 que houve uma expressão mais acentuada dos *no hits* na biblioteca de CC-Frutos est. tardio, além disso os *no hits* mais expressos GSCOCT00036640001, GSCOCT00002344001, GSCOCT00024650001 e GSCOCT00014971001, por exemplo, parecem estar presentes exclusivamente no estágio de desenvolvimento final do fruto de *C. canephora*, quando comparados aos demais estádios, e se encontram ausentes também no pericarpo, podendo ser exclusivos de perisperma e/ou endosperma.

Outros genes também foram específicos ao estágio de desenvolvimento do fruto, como no caso do GSCOCT00042747001, GSCOCT00019488001, GSCOCT00033656001, GSCOCT00018112001 e o GSCOCT00004205001, por exemplo. Estes *no hits* apresentaram expressão somente no estágio inicial de desenvolvimento e no pericarpo. No estágio intermediário de desenvolvimento do fruto (cerca de 30 semanas após a floração) a expressão do *contig* GSCOCT00007128001 é mais alta quando comparado com os demais neste estágio. Principalmente os *no hits* mais expressos neste estágio, GSCOCT00019370001, GSCOCT00004137001, GSCOCT00025611001, GSCOCT00027918001 e GSCOCT00028738001 foram exclusivos para o estágio intermediário de desenvolvimento. O *no hit* mais expresso no pericarpo foi o GSCOCT00042747001, no entanto alguns *no hits* foram específicos para este tecido, como por exemplo, o GSCOCT00042065001, GSCOCT00007086001 e o GSCOCT00002825001.

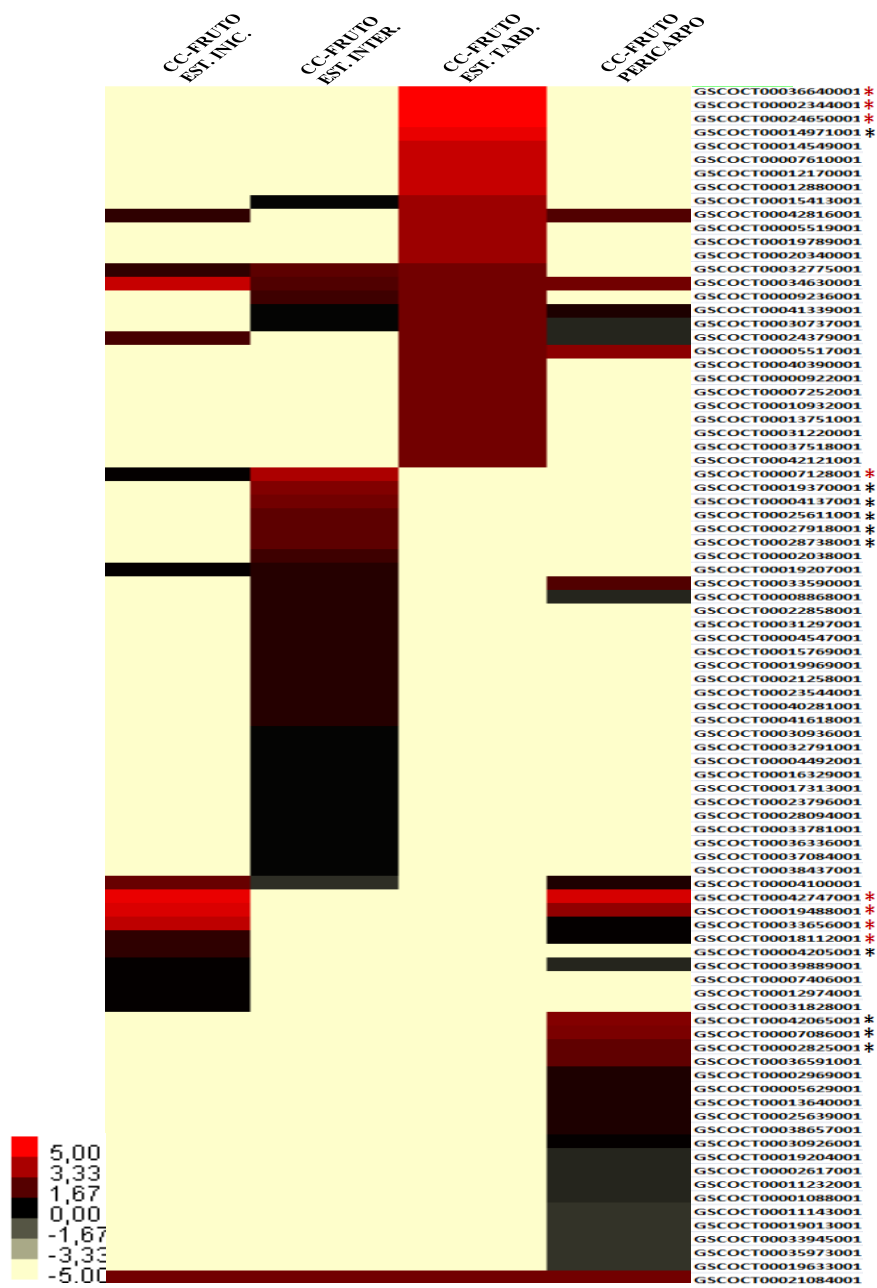


Figura 5 Expressão gênica dos *no hits* nas bibliotecas de frutos de *C. canephora* (CC) em diferentes estádios de desenvolvimento, utilizando a tecnologia Sanger. (*) *no hits* citados no texto. (*) *no hits* citados no texto e selecionados para qPCR.

3.3 Análise de expressão gênica *in silico* das bibliotecas de raiz de *C. canephora* (CC) utilizando a tecnologia 454

O resultado do *Northern eletrônico* dos *no hits* expressos em raiz de *Coffea canephora* (CC) utilizando a tecnologia de sequenciamento do tipo 454 podem ser visualizados na Figura 6. É notável perceber perfis de expressão diferencial entres os clones 14, 22, 73 e 120 e também perfis de expressão diferencial com relação às diferentes condições de estresse hídrico, sendo I a condição irrigado e NI a condição não-irrigado dentro dos clones.

O *no hit* mais expresso nas bibliotecas de raiz foi o GSCOCT00033656001. Este apresentou uma indução de expressão na condição de seca nos clones 22, 14 e 120, sendo que no clone 73 a expressão deste *no hit* foi praticamente a mesma para as duas condições analisadas. Alguns *contigs* apresentaram o mesmo comportamento de indução com o estresse de seca nos clones 22, 14, 73 e 120 e expressão igual para as duas condições no clone 73, esse quadro é destacado, principalmente, para o GSCOCT00005517001, GSCOCT00042816001, GSCOCT00019488001, GSCOCT00024379001, GSCOCT00007739001, GSCOCT00040390001, GSCOCT00005128001, GSCOCT00014149001 e GSCOCT00001872001.

Alguns *no hits* apresentaram perfis de indução com o estresse de seca em todos os clones analisados, como no caso, principalmente, do GSCOCT00034630001, GSCOCT00041471001, GSCOCT00041294001, GSCOCT00014549001, GSCOCT00041056001, GSCOCT00002617001 e GSCOCT00030106001. Este perfil de expressão foi relatado em Marraccini et al. (2012) para o também *no hits* *CcUNK10*. No entanto, outros *no hits* parecem ser reprimidos na condição de estresse de seca, como no caso do GSCOCT00003022001 e GSCOCT000017768001, que apresentaram diminuição da expressão na condição não irrigada em todos os clones

analisados. O *no hit* GSCOCT00041339001 também apresentou perfil de repressão com o estresse de maneira menos acentuada que os anteriores. É importante ressaltar *no hits* que apresentaram perfis opostos com relação à expressão nas diferentes condições, por exemplo, o *no hit* GSCOCT00004100001 teve um acentuado decréscimo de expressão na condição não irrigação do clone 22, porém no clone 14 este *no hit* se comportou de maneira inversa, no clone 73 ele não se expressou e no clone 120 este *no hit* apresentou uma pequena diminuição de expressão na condição de estresse.

O *no hit* GSCOCT00031831001 apresentou perfis de repressão para a condição de estresse de seca nos clones 22, 14 e 73, entretanto, no clone 120 este *no hit* parece ser induzido com o estresse.

Foi possível selecionar por meio desta análise *in silico* alguns genes *no hits* candidatos de raiz para os ensaios de qPCR, foram eles: GSCOCT00004100001; GSCOCT00014149001; GSCOCT00014549001, GSCOCT00034630001, GSCOCT00042816001 e GSCOCT00005517001.

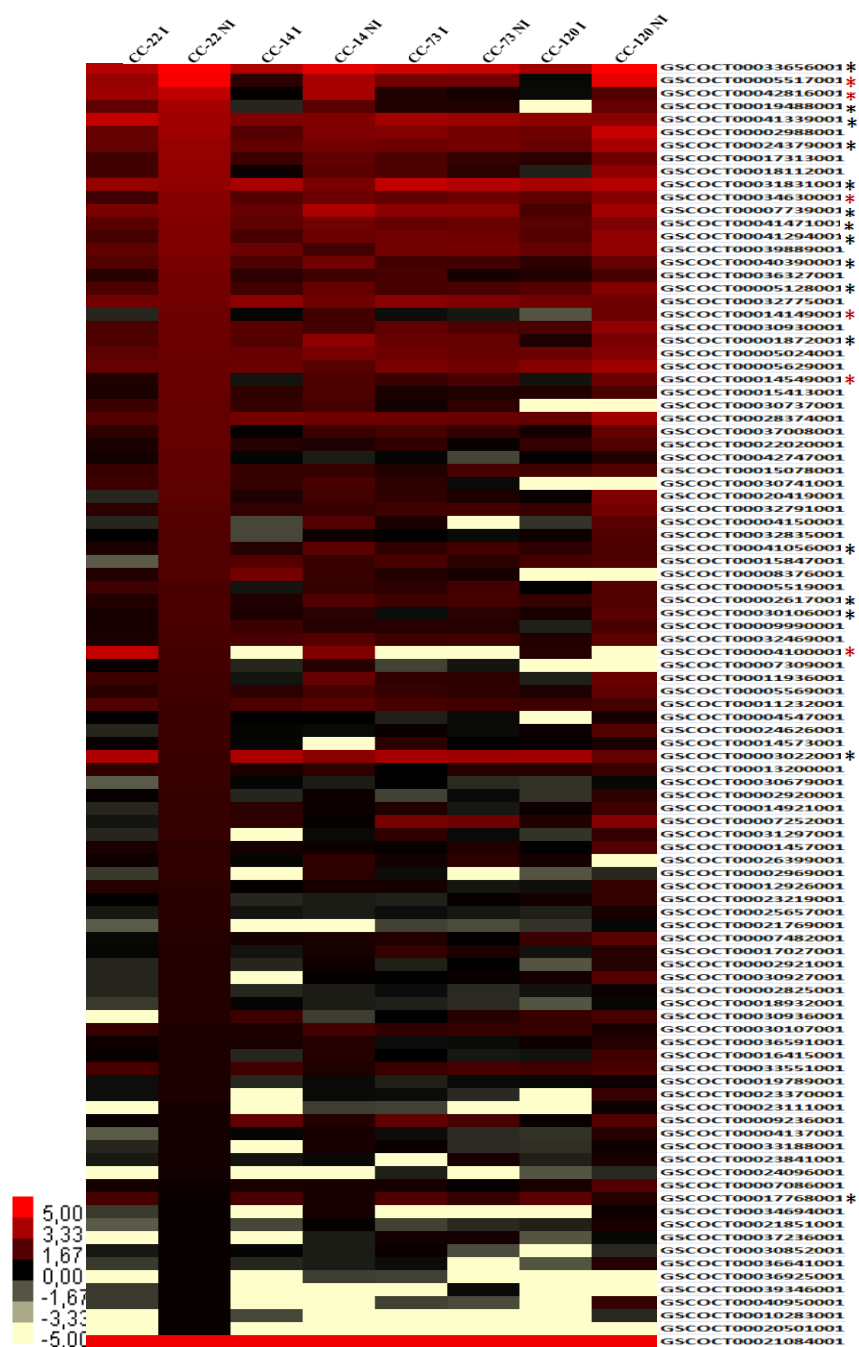


Figura 6 Expressão gênica dos *no hits* nas bibliotecas de raiz de *C. canephora* (CC) utilizando a tecnologia 454. (*) *no hits* citados no texto. (*) *no hits* citados no texto e selecionados para qPCR.

3.4 Análise de expressão gênica *in silico* nas bibliotecas de meristema de *C. arabica* (CA) utilizando a tecnologia 454

A Figura 7 apresenta o resultado do *nortern eletrônico* dos *no hits* expressos em meristema das cultivares Rubi (sensível à seca) e I59 (tolerante à seca) de *C. arabica* (CA) utilizando a tecnologia de sequenciamento do tipo 454. O perfil de expressão gênica parece ser diferencial para as cultivares analisadas e entre as condições de irrigação e não irrigação.

O *no hit* mais expresso nas cultivares foi o GSCOCT00017371001. Alguns *no hits* apresentaram perfis de indução com o estresse de seca nas duas cultivares analisadas, como no caso principalmente do GSCOCT00042816001, GSCOCT00033656001, GSCOCT00033590001, GSCOCT00034630001, GSCOCT00034721001, GSCOCT00015847001, GSCOCT00005128001, GSCOCT00025639001, GSCOCT00013200001, GSCOCT00004225001, GSCOCT00005569001, GSCOCT00007309001 e GSCOCT00007406001. No entanto, outros *no hits* parecem ser reprimidos na condição de estresse de seca nas duas cultivares analisadas, como no caso do GSCOCT00016329001, GSCOCT00015078001, GSCOCT00032791001, GSCOCT00037518001 e GSCOCT00036591001.

Outros genes apresentaram perfil de expressão somente na cultivar IAPAR59 na condição de estresse hídrico, como no caso do GSCOCT00017768001, GSCOCT00006372001, GSCOCT00020340001, GSCOCT00041957001 e GSCOCT00005005001.

No entanto, alguns genes apresentaram expressão somente na cultivar sensível à seca, dentre eles podemos citar o GSCOCT00011524001, GSCOCT00016725001, GSCOCT00040397001, GSCOCT00041294001 e GSCOCT00042115001, presentes na parte inferior da Figura 7. Contudo alguns *no hits* apresentaram expressão somente na cultivar sensível à seca na condição

de estresse hídrico, como exemplo o GSCOCT00012442001, GSCOCT00035511001, GSCOCT00007715001, GSCOCT00011862001 e GSCOCT00017847001. Estes *no hits* podem ser genes com indicativo de sensibilidade à seca para estas cultivares de *C. arabica*. Entretanto, houve *no hits* que parecem ser expressos somente na cultivar IAPAR59, como é possível observar nos perfis de expressão do GSCOCT00020419001, GSCOCT00035778001, GSCOCT00033240001 e GSCOCT00007252001. Estes *no hits* podem ser genes implicados na tolerância ao estresse de seca.

É importante ressaltar *no hits* que apresentaram perfis opostos com relação à expressão nas diferentes condições de irrigação e não irrigação, por exemplo o GSCOCT00030926001, GSCOCT00038657001, GSCOCT00025138001, GSCOCT00019488001, GSCOCT00041339001, GSCOCT00031297001, GSCOCT00030106001, GSCOCT00034759001, estes *no hits* parecem ser reprimidos com o estresse de seca na cultivar Rubi, se comportando de maneira oposta na cultivar tolerante I59. No entanto, outros *no hits* se comportam de maneira oposta, porém na cultivar sensível à seca, Rubi, estes genes parecem estar induzidos com o estresse hídrico, se comportando de maneira oposta ao tolerante à seca, I59, diminuindo a sua expressão na condição de estresse. Estes genes são o GSCOCT00035378001, GSCOCT00029497001, GSCOCT00019204001, GSCOCT00007482001, GSCOCT00009990001 e GSCOCT00028775001, visualizados na parte inferior da Figura 7.

Foi possível selecionar por meio desta análise *in silico* alguns *no hits* candidatos de meristema para os ensaios de qPCR, foram eles: GSCOCT00011524001 expresso somente na cultivar sensível à seca; GSCOCT00017768001 expresso somente na cultivar tolerante à seca na condição de estresse hídrico; GSCOCT00025639001 induzido com o estresse nas duas cultivares e GSCOCT00032791001, que parece ser reprimido com o estresse nas duas cultivares analisadas.

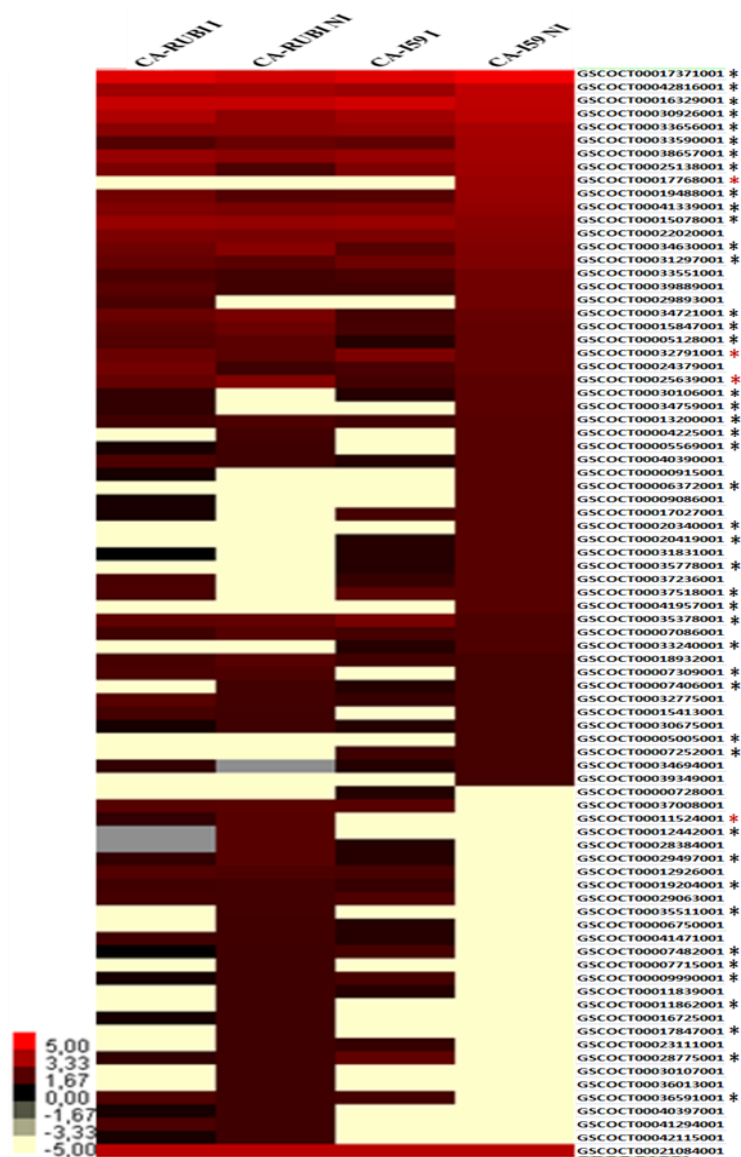


Figura 7 Expressão gênica dos *no hits* nas bibliotecas de meristemas de *C. arabica* (CA) utilizando a tecnologia 454. (*) *no hits* citados no texto. (*) *no hits* citados no texto e selecionados para qPCR.

4 CONCLUSÃO

Os genes órfãos (*no hits*) foram identificados e caracterizados *in silico* de acordo com o tecido em que foram expressos nas espécies *C. arabica* e *C. canephora*. Foi possível identificar e caracterizar *no hits* que parecem responder de forma significativa quanto à tolerância e suscetibilidade ao estresse de seca, assim como *no hits* que podem estar implicados na resposta das plantas a estresses bióticos ou lesões. As análises *in silico* apontam também *no hits* que são expressos diferencialmente nos tecidos e diferentes estádios de desenvolvimento do fruto de *C. canephora*. Alguns *no hits* que foram específicos para as bibliotecas de CA-Plântulas, CA-Callus, CA-Frutos, CC-Callus e CC-Frutos, podendo indicar possíveis promotores para novos processos biotecnológicos com tecido-especificidade.

REFERÊNCIAS

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**, Washington, v. 95, n. 25, p. 14863-14868, Dec. 1998.

HORAN, K. et al. Annotating genes of known and unknown function by large-scale co expression analysis. **Plant Physiology**, Bethesda, v. 147, n. 1, p. 41-57, 2008.

KEARSE, M. et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, Oxford, v. 28, n. 12, p. 1647-1649, Dec. 2012.

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**, Berlin, v. 112, n. 1, p. 114-130, Dec. 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**, Oxford, v. 63, n. 11, p. 4191-4212, 2012.

_____. *RBCSI* expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. **BMC Plant Biology**, London, v. 11, n. 85, p. 1-23, May 2011.

NAGARAJ, S. H.; GASSER, R. B.; RANGANATHAN, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. **Briefings in Bioinformatics**, London, v. 8, n. 1, p. 6-21, Jan. 2007.

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**, Berlin, v. 276, n. 5, p. 436-449, 2006. 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, Washington, v. 74, n. 12, p. 5463-5467, 1977.

TAUTZ, D.; DOMAZET-LOSO, T. The evolutionary origin of orphan genes. **Nature Reviews Genetics**, London, v. 12, n. 10, p. 692-702, Oct. 2011.

VIDAL, R. et al. High-throughput sequencing of cDNA shows that cv. Rubi and IAPAR59 of *Coffea arabica* have different molecular response to water privation. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 24., 2012, San José. **Proceedings...** San José: ICCS, 2012. 1 CD-ROM.

VIDAL, R. O. et al. A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. **Plant Physiology**, Bethesda, v. 154, n. 3, p. 1053-1066, 2010.

VIEIRA, L. G. E. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, Piracicaba, v. 18, n. 1, p. 95-108, Mar. 2006.

WINCKER, P. et al. Sequencing the coffee genome. In: PLANT AND ANIMAL GENOME CONFERENCE, 19., 2011, San Diego. **Proceedings...** San Diego: PAGC, 2011. p. 15-19.

XIE, C. et al. Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. **PLoS Genetics**, San Francisco, v. 8, n. 9, Sept. 2012. Disponível em: <<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002942>>. Acesso em: 5 fev. 2013.

ZHANG, W. Y. et al. A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. **PLOS One**, San Francisco, v. 6, n. 3, Mar. 2011. Disponível em: <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0017915>>. Acesso em: 5 fev. 2013.

CAPÍTULO 3

VALIDAÇÃO DOS GENES CANDIDATOS POR MEIO DA ANÁLISE DA EXPRESSÃO GÊNICA POR PCR QUANTITATIVO EM TEMPO REAL (qPCR)

1 INTRODUÇÃO

A capacidade da planta em manter suas atividades metabólicas e se adaptar enquanto se encontra estressada é mediada por mecanismos de respostas moleculares e bioquímicas, sofrendo então uma série de alterações morfológicas e fisiológicas. Em nível molecular, a expressão diferencial de genes para tolerar as condições adversas é um bom indicativo da resposta da planta ao meio ambiente.

Os mecanismos de resposta molecular ao estresse biótico e abiótico são controlados em função do tempo de duração e da intensidade do estresse e processam-se nos mais variados níveis de complexidade: morfológicamente, no caso do estresse de seca, por meio da redução da área foliar e do aumento do sistema radicular (volume e/ou profundidade); fisiologicamente, por meio de estratégias como o fechamento estomático, o ajuste osmótico, maior eficiência no uso da água e molecularmente, pela expressão diferencial de genes, conferindo ao vegetal o potencial de tolerar as condições desfavoráveis (PINHEIRO et al., 2005).

No caso de um estresse biótico como a infestação por nematoide na raiz, por exemplo, as células das raízes que cercam os nematoide apresentam características morfológicas e fisiológicas tais como acúmulo de compostos fenólicos e um aspecto de células necróticas, levando a danificação do tecido (ALBUQUERQUE et al., 2010). Em nível molecular, também há uma expressão diferencial de genes para tentar tolerar essa condição de estresse.

O sabor e aroma do café também são os produtos de um complexa cadeia de transformações bioquímicas. Um dos principais objetivos da pesquisa sobre a qualidade do café é o de estabelecer relações entre a composição química dos grãos de café e a qualidade da bebida, e em particular, para

identificar os precursores gênicos envolvidos no desenvolvimento do aroma e sabor.

Sendo assim, as pesquisas em biotecnologia estão trazendo grande impacto para a agricultura e a indústria alimentícia em todo o mundo. Onde a utilização destas tecnologias aplicadas à análise da expressão de genes têm sido de fundamental importância para o desenvolvimento científico e elucidação de processos moleculares.

Objetivou-se com as pesquisas realizadas para este capítulo foi validar a expressão de 17 genes candidatos selecionados a partir das análises *in silico*, por meio da técnica de PCR quantitativo em tempo real (qPCR), elucidando sobre a presença e expressão dos *no hits in vivo* nos diferentes tecidos dos materiais vegetais de *C. arabica* e *C. canephora* submetidos a condições biológicas de estresse biótico e abiótico.

2 MATERIAL E MÉTODOS

2.1 Material Vegetal

O material vegetal selecionado para as análises de expressão gênica em qPCR são provenientes de diferentes tecidos (folhas, raízes, frutos e meristemas) das espécies de *C. arabica* e *C. canephora*, submetidos a diferentes condições biológicas (estresse biótico e abiótico) e frutos em diversos estádios de desenvolvimento (Tabela 1). Os métodos de cultivo e colheita de alguns materiais utilizados já foram publicados, os demais serão mais detalhados a seguir.

Tabela 1 Material vegetal utilizado para a extração de RNA e condução dos experimentos de qPCR.

Espécie	Origem	Tecido	Condição de estresse biótico	Condição de estresse abiótico	Estádios de desenvolvimento	Publicação
<i>C. arabica</i>	Rubi e IAPAR59	Meristemas		Irrigado e Não-irrigado	Maduros	Vidal et al. (2012)
<i>C. canephora</i>	Clones 14, 22, 73 e 120	Folhas		Irrigado e Não-irrigado	Folhas adultas	Marraccini et al. (2012)
<i>C. canephora</i>	Clones 14, 22, 73 e 120	Raiz		Irrigado e Não-irrigado	Raiz desenvolvida	Não publicado
<i>C. canephora</i>	Clones 14 e 22	Raiz	Inoculado com <i>Meloidogyn</i> e <i>Incognita</i>		Raiz desenvolvida	Não publicado
<i>C. canephora</i>	Conilon ¹	Frutos-perisperma, endosperma e pericarpo			60, 90, 120, 150, 180, 210, 270 e 300DAF*	Não publicado

¹ Planta cultivada no campo –CPAC- linha 6, planta 35

* Dias Após a Floração

2.1.1 Meristemas de *C. arabica* cultivados no campo

Foram utilizadas para a coleta dos meristemas, plantas oriundas das cultivares de *C. arabica* que apresentam comportamento diferente vista à seca. A cultivar Rubi considerada sensível à seca e a cultivar Iapar59 (I59) considerado tolerante à seca (VIDAL et al., 2012). Para analisar as respostas de essas cultivares à seca, as plantas foram cultivadas no campo experimental da Embrapa Cerrados – CPAC (Planaltina-DF, Brasil) em duas condições experimentais: com irrigação (I, plantas controle não estressadas) e sem irrigação (NI, plantas estressadas) durante o período de seca (inverno). Os ápices foram coletados a partir da extremidade dos ramos plagiotrópicos na estação de seca (Agosto, 2011). O estresse foi avaliado medindo os potenciais hídricos foliares de antemanhã (Ψ_{am}) das plantas dos cultivares nas condições irrigado (I) e não irrigado (NI). Para os dois cultivares, os valores do Ψ_{am} foram de -0,2 e <-3,5 MPa para as condições I e NI, respectivamente. Após, os ápices foram conservados a -80 °C.

2.1.2 Folhas e raízes de *C. canephora* dos clones 14, 22, 73 e 120 cultivados em casa de vegetação

Foram utilizadas folhas e raízes dos clones de *C. canephora* var. *conilon* tolerante (14, 73 e 120) e suscetível (22) à seca, vindos do Instituto de Pesquisa e Assistência Rural (Incaper, Vitória, Espírito Santo, Brasil). Estes clones foram cultivados individualmente em vasos de 12 litros em casa de vegetação na UFV (Universidade Federal de Viçosa-UFV, Minas Gerais, Brasil), em condições controladas de temperatura (25 °C) e umidade (70%). Após 6 meses, as mudas de cada clone foram separados em dois grupos: o primeiro (I) recebeu irrigação regular (controle), enquanto a irrigação foi suspensa atingido um Ψ_{am} de -3,0

MPa. Para ambas as condições, 6 plantas (repetições biológicas) foram analisadas. Foram utilizadas para as medidas somente folhas totalmente expandidas e não danificadas, localizadas no terceiro par a partir da extremidade dos ramos plagiotrópicos. As folhas foram coletadas às 10 h e imediatamente conservadas em nitrogênio líquido e depois estocadas a -80 °C (MARRACCINI et al., 2012). No caso das raízes, estas foram coletadas, lavadas e posteriormente armazenadas a -80 °C.

2.1.3 Raízes de *C. canephora* inoculadas com o nematoide *Meloidogyne incognita*

Os materiais vegetais utilizados para inoculação são provenientes de raízes de *C. canephora* do clone 22 (considerado suscetível ao nematoide *Meloidogyne incognita*) e do clone 14 (identificado como resistente a *M. incognita*) oriundos do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) (CARNEIRO et al., 2009). Estacas dos diferentes clones foram plantadas em areia esterilizada na Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF. Ficaram aclimatando até o inóculo ficar pronto. Os nematoide foram reproduzidos em tomateiro cv. Santa Cruz, na EMBRAPA, após 3 meses os tomateiros foram retirados dos vasos e a parte aérea foi descartada. As raízes, então, foram lavadas, cortadas e batidas no liquidificador com hipoclorito de sódio a 0,5% de acordo com o método de Hussey e Baker (1973), modificado por Boneti e Ferraz (1987). Em seguida, as raízes batidas foram colocadas em conjunto de peneiras de 20, 100 e 500 mesh, respectivamente. Na peneira de 500 mesh ficam retidos os ovos. Após a lavagem para retirada do hipoclorito, o extrato de ovos foi retirado e tratado de acordo com o método de Flegg (1967), para a obtenção dos juvenis, aqui descrito: A solução de ovos foi posta em funil de Baermann e mantida em água destilada

para eclosão dos juvenis. Uma vez liberado na água, esta foi reduzida por meio de sifão e centrifugação. Os juvenis foram então contados com o auxílio de câmara de Peters e microscópio de luz a fim de se calcular quantos juvenis haviam por ml. Foram inoculados 10.000 juvenis por planta de café nos clones 14 e 22. As raízes foram então coletadas no 4º, 8º, 12º, 20º, 32º e 45º dias após a inoculação para os dois clones e armazenadas a -80 °C para posterior extração de RNA. Foi retirada, também, uma raiz de uma planta não inoculada de cada genótipo de café testado, como o controle para a expressão gênica. Neste trabalho foi feito um mix com os cDNAs no 4º, 8º, 12º e 20º dia após a inoculação para analisar a expressão dos genes nos clones 14 e 22, com e sem a presença do nematoide.

2.1.4 Frutos de *C. canephora* cultivados no campo

Frutos utilizados para a análise são provenientes de um clone (linha 6, planta 35) de *C. canephora* cultivado no campo experimental da Embrapa Cerrados – CPAC (Planaltina-DF, Brasil). Esta planta foi escolhida, pois apresenta padrões de características positivas em relação ao fenótipo como o alto vigor, onde é levado em consideração a quantidade de folhas que a planta possui e a quantidade de ramos depois da frutificação que fornecem uma estimativa de produtividade. Ela apresentou uma carga alta de produtividade dos frutos, apresentou também um padrão de maturação dos frutos tardio em relação aos demais. A coleta foi realizada mensalmente no período de novembro de 2011 à Julho de 2012. A primeira coleta dos frutos foi realizada em novembro de 2011, sendo 60 dias após a floração (DAF) deste ano, em seguida foram coletados os períodos de 90 DAF, 120 DAF, 150 DAF, 180 DAF, 210 DAF, 270 DAF e 300DAF.

2.2 Extração de RNA e quantificação das alíquotas

Todos os materiais utilizados para extração (cadinhos, pistilos de porcelana, microtubos, ponteiras e água destilada) foram autoclavados para limitar a ausência de atividade da RNase e garantir a qualidade do RNAs extraídos. Os RNAs das folhas, frutos, meristemas e raízes inoculadas com nematoide, descritos anteriormente, foram extraídos utilizando o protocolo de extração de RNA com o reagente Concert, realizado de acordo com manual: Concert™ Plant RNA Purification Reagent (Invitrogen). Os RNAs de raiz I e NI foram extraídos utilizando o protocolo de extração de RNA com o reagente Trizol® Reagent.

Para se avaliar a integridade das amostras extraídas, o RNA foi submetido à eletroforese em gel de agarose 1,0% corado com Brometo de Etídio ($0.5 \mu\text{g. mL}^{-1}$), posteriormente visualizado sob luz ultravioleta. A imagem foi captada pelo foto-documentador da Loccus Biotecnologia, podendo ser visualizadas as bandas individuais de RNA ribossomal 26S e 18S. As amostras foram quantificadas em espectrofotômetro (Nanodrop® Espectrophotometer ND-1000). A qualidade foi avaliada pelo espectro (220-600nm) com a razão OD_{260}/OD_{280} . Se a razão for menor que 1,8, pode indicar a presença de contaminação de proteínas nas amostras de RNA.

2.3 Tratamento com DNase

Todas as amostras foram tratadas com o Kit RQ1 RNase-Free DNase (Promega), de acordo com o protocolo do fabricante, para retirada dos vestígios de DNA genômico que poderiam sobrar nas amostras de RNA. As reações foram realizadas utilizando o termociclador modelo T-100 (BioRad). Ao microtubo contendo 8 μg de RNA total foram adicionados 1 μL de tampão 10x

(presente no kit) e 1 μL da enzima RQ1 RNase Free. Essa reação foi incubada a 37 °C por 30 min. para ativação da enzima. Para inativar a enzima foi adicionado 1 μL de RQ1 RNase Stop Solution (EGTA 20 mM pH 8) e a reação foi incubada aos 65 °C por 10 min. Para a avaliação da integridade e da qualidade das amostras, o RNA tratado foi submetido à eletroforese em gel de agarose e as amostras foram quantificadas.

2.4 Transcriptase reversa (RT)

Depois de verificada a ausência de gDNA nas amostras, foi realizada a síntese reversa da primeira fita do DNA complementar (cDNA) a partir dos RNAm utilizando o *primer* oligo dT₁₅, o kit ImProm™ Reverse Transcriptase (Promega) e o termociclador modelo T-100 (BioRad). Em microtubos estéreis, adicionou-se 1 μg de RNA total, 1 μL de oligo- dT₁₅ concentração final de 10 μM . A reação foi incubada a 70 °C por 5 minutos e rapidamente levada a 4 °C por 5 minutos, para desnaturação das estruturas secundárias e para o anelamento dos primers ao RNA. Após serem mantidas a 4 °C por 5 min acrescentou-se ao mesmo tubo 4 μL do tampão ImProm™ 5x Reaction Buffer, 1 μL de RNase Out (40 U. μL^{-1}), 2,4 μL de MgCl₂ na concentração de 25 μM , 1 μL do mix de dNTP na concentração final de 0,5 mM cada dNTP e 1 μL da enzima ImProm™ Reverse Transcriptase e água Milli-Q para se completar o volume final de 20 μL . Os tubos foram incubados no termociclador, a 25 °C, durante 5 min para a fase de anelamento. A reação foi incubada aos 42 °C por 60 minutos para a síntese da primeira fita de cDNA, sendo posteriormente incubada a 70 °C durante 15 minutos para inativar a enzima. As amostras de cDNA foram armazenadas a -20 °C.

2.5 Desenho dos *primers* para as análises de expressão por qPCR

Os *primers* requeridos para se analisar a expressão dos *no hits* foram desenhados utilizando o programa “Primer Express 3.0” (Applied Biosystems). Eles foram nomeados de acordo com a numeração encontrada nas análises *in silico* no Capítulo 2, retirando o prefixo GSCOCT000 e o sufixo 001. (Tabela 3) Os estudos de quantificação relativa por PCR em tempo real exigem conjuntos de *primers* com condições específicas, sendo assim, o uso de *primers* utilizados em ensaios de PCR tradicional nem sempre é recomendado para qPCR. Para a técnica de qPCR o comprimento do amplicon deve ser menor que 200 pb e a temperatura melting (T_m) (temperatura de desnaturação de 50% das moléculas) deve ser próximo de 60 °C (maior que a T_m da PCR convencional). Além disso, para a análise por meio do método do C_T comparativo, a eficiência da amplificação do alvo (gene de interesse) e a eficiência da amplificação da referência (controle endógeno) devem ser aproximadamente iguais. O *primer* utilizado como controle endógeno para todas as reações de qPCR foi o NUBI, referente ao gene da Ubiquitina10 (*CcUBQ10*).

Tabela 3 *Primers forward (F) e reverse (R) utilizados para as análises de qPCR. NUBI corresponde ao gene endógeno CcUBQ10. Os primers estão identificados de acordo com a nomenclatura dos no hits encontrados nas análises in silico, sem o prefixo GSCOCT000 e o sufixo 001.*

<i>Nome dos Primers</i>	<i>Sequências dos Primers (5'-3')</i>	<i>Amplicon (pb)</i>
NUBI-F	AAGACAGCTTCAACAGAGTACAGCAT	104
NUBI-R	GGCAGGACCTTGGCTGACTATA	
14149-F	CTCCCTTCCGGGTACCGATA	80
14149-R	AGCCCGGAGAGCTTCTTTTG	
14549-F	TCAACGCCAGTTCGAATTC	85
14549-R	GTGGTGGCGCGTCTCCTA	
34630-F	CGGCTTACCGAAGAAATCA	61
34639-R	GGTGGGTAAGTGCCTTGTT	
42816-F	GCCTGAAAGAGGTCCAACCA	78
42816-R	CTCCGAACCTTGAAAACAGAAGA	
33656-F	CTATGGTGGTGGTGCTGGCTA	110
33656-R	CTTACCAAGGTGCTCAAGATTT	
30926-F	AGGCTACGGAGGAGGCCAT	123
30926-R	CCGTGTCCATAACCACCATGT	
11524-F	GGCGTTGTCATGGATACG	119
11524-R	TTTGGCTCACCAGCATATGTG	
17768-F	TCGTCTTGCCAGCATCTCA	60
17768-R	CGGACAATCAGCGTCATGTG	
25639-F	CGAGGAAGCTGAAGGAAAGGA	61
25639-R	TCCGACTGGCCTAACAAGGT	
32791-F	CCGAGCACCATTTCCAACAC	102
32791-R	CCTTCTTCCTTTCTGGCTTCAA	
18112-F	TGTGAACTGCCATCCCAAGA	88
18112-R	AAGACTACCATGTCCAACAACCTCAG	
07128-F	TCCTGAACTGCCATCACCAA	80
07128-R	GGCCATTGAACACCCCATAC	
05517-F	CGGATTCACGTGGAAATGTGAG	130
05517-R	CCAGAGGAGTTTGGTTTCACCC	
02344-F	TCAATCGCCATCTGCTAACG	120
02344-R	CTCCGAGACGCAGGAATCTC	
24645-F	AAGACCCCACTCGACTTTTGAC	79
24645-R	TCTGCAACGCGTTTCTTTTG	
22858-F	ATCAAGCGTAGCGCAATGC	103
22858-R	ATAGCATAGGCAGGCGCATC	
33590-F	CACCACCACCTCCAGTGCATAT	108
33590-R	GTGGTGGTGGTACTTATACTTGAAA	

2.6 Avaliação da eficiência dos *primers*

Os pares de *primers* usados para as análises de qPCR são descritos na Tabela 3. Para cada par de *primers*, foi realizado um ensaio prévio de quantificação absoluta, para determinar a eficiência e a diluição mais adequada de cDNA das amostras que foram usadas. Os cDNAs foram diluídos 1:10, 1:25 e 1:50, e testados com o gene controle endógeno *CcUBQ10*. Após determinar a melhor diluição do cDNA, foram realizadas quatro diluições seriadas do concentrado de cDNA (120 ng. μL^{-1}) (fator de diluição de 1/10). Todos os *primers* testados apresentaram uma boa eficiência de amplificação (*amplicon* da PCR dobra de quantidade durante a fase linear da amplificação de PCR) cujos valores de eficiência estão entre 90-100%, encontrando-se próximos daqueles aceitáveis ($100 \pm 10\%$) (manual da Applied Biosystems).

2.7 Análises da expressão por PCR quantitativo em tempo real

Para cada reação, foi utilizado 1 μL das diluições adequadas de cDNA, 0.2 μL do *primer forward*, 0.2 μL do *primer reverse* a 10 mM cada e 5 μL de Platinum[®] SYBR[®] Green qPCR SuperMix-UDG with ROX (Invitrogen), este fluoróforo tem a capacidade de intercalar nas fitas duplas de cDNA. A reação foi completada com 3.6 μL de água Milli-Q para um volume final de 10 μL por reação. Sendo que para cada condição foi feita a mesma reação três vezes para sobrepor e confirmar os resultados no aparelho. Os dados foram analisados no programa 7500 Fast Software (software v2. 0.1). As amostras foram processadas em triplicatas, sempre acompanhadas dos controles negativo (NTC: “*no template control*”) que não contém cDNA. O controle negativo nas reações é usado para verificar a ausência de contaminação de cDNA exógeno nos mixes de SYBR mais *primers* ou na água. Com o objetivo de se verificar a especificidade de

anelamento dos *primers* aos fragmentos alvos de interesse, foi realizada a análise das curvas de dissociação dos fragmentos amplificados ao final dos ensaios de qPCR, para cada par de *primers* utilizado. A expressão dos genes alvo foi normalizada com o gene *CcUBQ10*, considerado controle endógeno eficiente para café (BARSALOBRES-CAVALLARI et al., 2009). A normalização foi realizada utilizando-se a equação $\Delta C_T = C_T (\text{gene alvo}) - C_T (\text{controle endógeno})$. A calibração foi determinada pela fórmula $\Delta\Delta C_T = \Delta C_T (\text{amostra}) - \Delta C_T (\text{calibrador})$. O calibrador é uma amostra usada como base para resultados de expressão comparativa. A quantificação relativa foi obtida pela fórmula $2^{-\Delta\Delta C_T}$. Os resultados foram normalizados usando C_{Ts} (Ciclo *Threshold*) obtidos para controles endógenos presentes na mesma reação. O C_T foi determinado pelo número de ciclos no qual a fluorescência gerada dentro de uma reação cruza o limiar (“*Threshold*”). O método usado foi o C_T comparativo (quantificação relativa).

3 RESULTADOS E DISCUSSÃO

Para as análises de qPCR foram selecionados 17 genes candidatos potencialmente envolvidos na resposta da planta aos estresses bióticos e abióticos, assim como genes implicados no desenvolvimento do fruto de café. Estes genes foram selecionados a partir das análises *in silico* apresentadas no Capítulo 2. Dentre eles, os genes *no hits* 33590, 33656 e 30926 foram avaliados quanto à expressão destes em raízes de *C. canephora* inoculadas com o nematoide *M. incognita*.

Os genes *no hits* 30926 e 33656 também foram testados em folhas de *C. canephora* de clones estressados e não estressados hidricamente. No entanto, os genes *no hits* 14149, 14549, 05517, 34690 e 42816 foram analisados quanto à expressão destes nas raízes de *C. canephora* de clones estressados e não estressados hidricamente. Outros genes *no hits* foram avaliados em meristemas dos cultivares Rubi e I59 de *C. arabica* com e sem estresse hídrico, foram eles: 11524, 17768, 25639, 30926 e 32791. Foram selecionados e testados também os genes *no hits* 07128, 18112, 33656, 02344 e 24645 em perisperma, endosperma e pericarpo de frutos de *C. canephora* em diferentes estádios de desenvolvimento. 33656 e 30926 foram analisados em mais de uma condição. (Tabela 4)

Tabela 4 *Primers* dos *no hits* utilizados nos diferentes materiais vegetais analisados nos ensaios de qPCR.

<i>No hits</i>	Tecidos/condição biológica
33590, 33656 e 30926	Raízes de <i>C. canephora</i> inoculadas com o nematoide <i>M. incognita</i> .
30926 e 33656	Folhas de <i>C. canephora</i> de clones estressados e não estressados hidricamente.
14149, 14549, 05517, 34690 e 42816	Raízes de <i>C. canephora</i> de clones estressados e não estressados hidricamente.
11524, 17768, 25639, 30926 e 32791	Meristemas de cultivares de <i>C. arabica</i> com e sem estresse hídrico.
07128, 18112, 33656, 02344 e 24645	Perisperma, endosperma e pericarpo de frutos de <i>C. canephora</i> em diferentes estádios de desenvolvimento.

Outros genes selecionados nas análises *in silico* tiveram os *primers* desenhados, no entanto eles não foram específicos ou não amplificaram como é possível observar na curva de dissociação apresentada na Figura 1. O *primer* 42747 apresentou três picos de dissociação para o mix de cDNAs dos estádios nos tecidos perisperma, endosperma e pericarpo.

É interessante ressaltar que para cada tecido este *primer* apresentou um TM (“*melt temperature*”) diferente, onde no perisperma (cor verde) a TM ficou em torno de 72 °C, no endosperma (cor azul claro), é possível observar um “ombro” em 69 °C e um pico em torno de 85 °C e no pericarpo (cor azul escuro), é possível observar um “ombro” em 69 °C e um pico em torno de 83,5 °C. Isto pode indicar a amplificação de formas alélicas diferente para este gene dentro desses diferentes tecidos do grão de café, e/ou até mesmo a ocorrência de um *splicing* alternativo para o mesmo gene, onde cada produto gênico é específico para determinado tecido.

Para testar essas hipóteses, outros experimentos com novos pares de *primers* para estes genes serão necessários. Para o *primer* 36640 (Figura 1B) é possível observar diferentes picos de TM para dissociação do *primer*

independente do tecido analisado mostrando não ser específico. Os *primers* 19488 e 04100 não amplificaram (dados não mostrados) e foram descartados para as análises. Estes deverão ser redesenhados e novas análises serão realizadas.

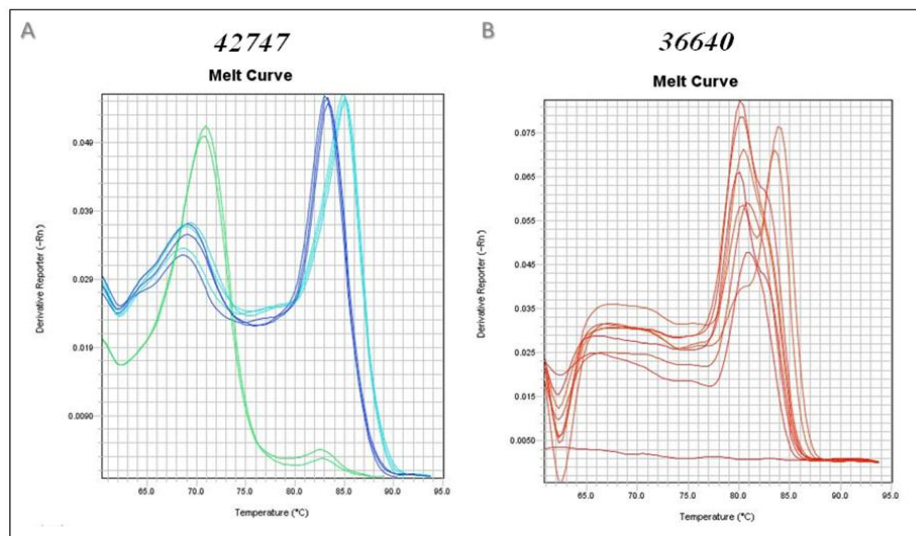


Figura 1 Curva de dissociação da qPCR para os *primers* 42747 (A) e 36640 (B) utilizando o mix dos estádios de desenvolvimento para os diferentes tecidos, perisperma, endosperma e pericarpo de frutos de *C. canephora*.

3.1 Expressão dos genes *no hits* potencialmente envolvidos na resposta da planta ao estresse biótico

Os clones analisados para a expressão dos genes candidatos na condição de estresse biótico foram o 14 e 22 de *C. canephora*. Sendo o clone 22 considerado sensível ao nematoide *Meloidogyne incognita* e o clone 14 considerado tolerante ao mesmo (CARNEIRO et al., 2009). Neste contexto, os genes 33590, 30926 e 33656 apresentaram perfis de expressão diferencial entre

os clones 14 e 22 submetidos às condições de estresse biótico (Figuras 2 A, B, C). Foi observado um aumento da expressão dos genes *33590* e *30926* nas raízes das plantas inoculadas com o nematoide *M.incognita* quando comparados com a expressão destes nas raízes não inoculadas. Este perfil foi observado nos dois clones analisados. Visto que para os dois genes em questão, a expressão nas raízes do clone 22 submetido ao estresse biótico parece ser maior do que a expressão destes genes no clone 14 submetido às mesmas condições.

Entretanto, o perfil de expressão do gene *33656* se comportou de maneira diferente. No clone 14, a sua expressão parece ser induzida com o estresse biótico enquanto esta expressão parece ser reduzida com o estresse no caso do clone 22. Além disso, a expressão do gene *33656* nas raízes do clone 22 não inoculadas com o nematoide parece ser 5x maior do que nas raízes não inoculadas do clone 14. Sendo que o clone tolerante 14 apresentou uma resposta de indução com o estresse biótico e o clone susceptível 22, apresentou perfil de expressão inverso, sendo reprimido com o estresse.

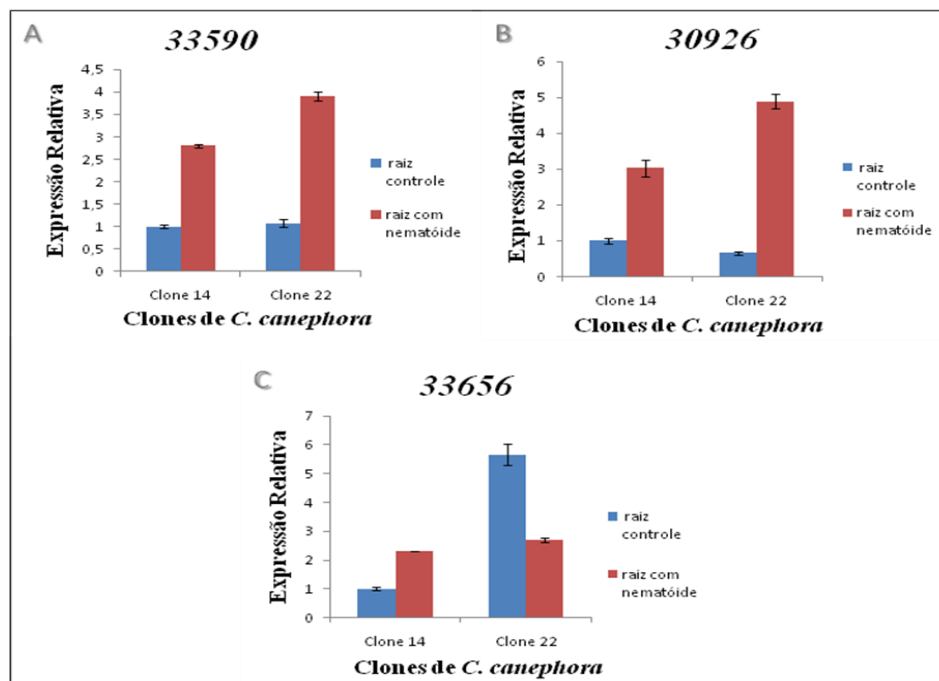


Figura 2 Perfis de expressão dos *no hits* 33590 (A), 30926 (B) e 33656 (C) nas raízes dos clones 14 e 22 de *C. canephora* submetidos (raiz com nematoide) ou não (raiz controle) ao estresse biótico. O calibrador interno utilizado foi o 14 controle (sem infecção de nematoide). O gene *CcUBQ10* foi utilizado como controle endógeno.

3.2 Efeito do estresse biótico sobre a expressão dos genes *no hits* em café

As análises dos perfis de expressão dos genes *no hits* selecionados mostram que a presença dos nematoide nas raízes das plantas afeta a expressão dos genes nos clones 14 e 22. Sendo que os genes 33590 e 30926 foram mais expressos no clone 22 sensível ao nematoide, quando este está presente, quando comparado ao clone tolerante 14. Isto pode indicar a participação destes genes no mecanismo de resposta das plantas à presença do patógeno, no caso o nematoide *M. incognita*, corroborando com os resultados apresentados no

Capítulo 2 quanto à possível participação destes genes na resposta da planta à patógenos ou às lesões.

3.3 Expressão dos genes *no hits* potencialmente envolvidos nos mecanismos de resposta à seca nas folhas de café

Os genes 33656 e 30926 foram selecionados para análise da expressão em folhas de clones de *C. canephora* tolerantes à seca 14, 73 e 120 e o clone 22 sensível à seca. Os clones analisados foram estressados (não irrigado, NI) e não-estressados (irrigado, I) hidricamente. Estes *no hits* foram selecionados a partir das análises *in silico* (Figura 4, Capítulo 2), e também já haviam sido selecionados para as análises de qPCR de raízes de clones de *C. canephora* com ou sem a presença do estresse biótico a partir das análises *in silico* (Figura 2, Capítulo 2). Estes genes podem estar implicados na cascata de resposta aos estresses bióticos e abióticos concomitantemente, participando de vias de transdução dos sinais relacionados aos estresses nas plantas.

3.4 Expressão do gene 33656 em folhas de *C. canephora* submetidos ao estresse de seca

As folhas das plantas dos clones 14, 22, 73 e 120 de *C. canephora* cultivadas com (I) ou sem (NI) irrigação, foram analisados quanto à expressão dos genes *no hits*. É possível observar perfis de expressão diferencial para o *no hit* 33656 entre os clones 14, 22, 73 e 120, e uma expressão diferencial também nas condições irrigadas ou de seca (Figura 3A). Este gene parece ser induzido com a seca nos clones 14, 22 e 120. No clone 73, o gene 33656 parece não apresentar variações de expressão significativa entre as duas condições analisadas (I e NI). O clone 22 apresentou uma expressão deste gene na

condição de seca, cerca de 60x maior do que a expressão na condição de irrigação neste mesmo clone. A expressão no clone 22NI parece ser 3x maior também do que no clone 14NI, 10x maior do que no clone 73NI e aproximadamente 6x mais do que no clone 120NI. Este perfil de expressão para o *no hit* 33656 em folhas dos clones de *C. canephora* já foi apresentado para o também *no hit* *CcUNK10* (MARRACCINI et al., 2012).

3.5 Expressão do gene 30926 em folhas de *C. canephora* submetidos ao estresse hídrico

O gene 30926 apresentou perfil de expressão diferencial entre os clones testados (Figura 3B). Assim, é possível observar uma indução da expressão do gene 30926, cerca de 5x maior na condição de estresse hídrico no clone 14, quando comparado ao mesmo clone sem estresse e cerca de 4.5x maior na condição de estresse hídrico no clone 120, quando comparado ao mesmo clone nas condições de irrigação. Não houve variação de expressão significativa da condição NI ou I do gene 30926 para os clones 22 e 73. Visto que, para este gene parece não haver oscilação considerável da expressão entre os clones na condição de irrigação (I), variando somente para condição de estresse hídrico (NI).

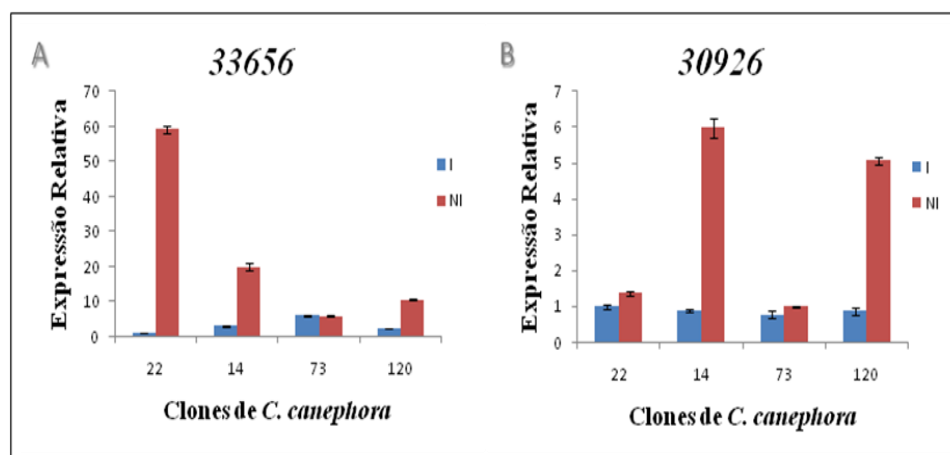


Figura 3 Perfis de expressão dos genes candidatos 33656 (A) e 30926 (B) em folhas dos clones 22, 14, 73 e 120 de *C. canephora* submetidos ao estresse de seca (NI: Não irrigado) e sem estresse (I: Irrigado). O calibrador interno utilizado foi o 22I. O gene *CcUBQ10* foi utilizado como controle endógeno.

Estes genes foram selecionados nas análises *in silico* por estarem presentes nas bibliotecas de folhas estressadas (CC-Folha estresse) e não estressadas (CC-Folha controle) com a seca (Capítulo 2). Mesmo havendo uma variação entre os clones os resultados apresentados para o *no hit* 30926 confirmam os resultados das análises *in silico*, sendo induzidos com o estresse de seca.

3.6 Expressão dos genes *no hits* potencialmente envolvidos no desenvolvimento do fruto de café

Com os dados apresentados nas análises *in silico* (Capítulo 2), foi possível selecionar *no hits* expressos nas bibliotecas de *C. canephora* de frutos no estágio inicial, intermediário e tardio de desenvolvimento como relatado em Lin et al. (2005). No entanto, as análises de expressão *in vivo* aqui apresentadas foram mais detalhadas quanto ao tecido e estágio de desenvolvimento em que os

no hits foram expressos. Enquanto que Lin *et al.* 2005 analisou os tecidos perisperma e endosperma juntos, para este trabalho os tecidos perisperma (PE), endosperma (END) e pericarpo (PO) foram analisados separadamente. Lin *et al.* (2005) também analisaram juntos os estádios iniciais de desenvolvimento do fruto de *C. canephora* (até 150DAF), os estádios intermediários de desenvolvimento (até 210DAF) e também os estádios tardios (até o 300 DAF) de desenvolvimento do fruto, além de analisarem o pericarpo de todos os estádios de desenvolvimento juntos. Para este trabalho, os estádios de desenvolvimento foram analisados separadamente para cada tecido (Figuras 4, 5, 6, 7 e 8).

Cinco *no hits* selecionados nas análises *in silico* apresentadas (Figura 5, Capítulo 2) foram testados quanto à expressão nos frutos de *C. canephora*. Primeiramente, foi analisada uma mistura de cDNAs dos perispermias (PE) de diferentes estádios, assim como uma mistura dos cDNAs de endospermias (END) nos diferentes estádios e, por fim, uma mistura do pericarpo (PO) nos diferentes estádios (Figuras 4A, 5A, 6A, 7A e 8A). A expressão dos *no hits* 33656, 07128, 18112, 02344 e 24645 foi realizada com essas misturas para conferir a qualidade (especificidade dos fragmentos amplificados) das reações de qPCR e posteriormente detalhada quanto ao estádio de desenvolvimento para cada tecido.

É possível observar nas Figuras 4, 5, 6, 7 e 8 que a separação dos tecidos foi bem executada, uma vez que não parece haver contaminação entre os tecidos estudados, pois para o mesmo gene e mesmo estádio de desenvolvimento do fruto há perfis diferenciais de expressão. Podemos observar ainda uma expressão diferencial dos *no hits* nos três tecidos analisados (PE, END e PO) e nos diferentes estádios de desenvolvimento analisados, estádio inicial (90DAF até 150 DAF), intermediário (180DAF e 210DAF) e tardio (270DAF e 300DAF).

3.7 Expressão do gene 33656 em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos

Os resultados das análises de qPCR mostraram que o gene 33656 (Figura 4) parece ter uma expressão maior no pericarpo (PO) quando comparado ao perisperma (PE) e endosperma (END). Com relação à expressão deste gene no perisperma (PE), é possível observar um aumento gradual da expressão ao longo do desenvolvimento do fruto (Figura 4C). Entretanto, no endosperma (END) a expressão do gene 33656 diminui ao longo do desenvolvimento (Figura 4D). No pericarpo (PO) há uma diminuição da expressão no estágio tardio de desenvolvimento de 300DAF (Figura 4E).

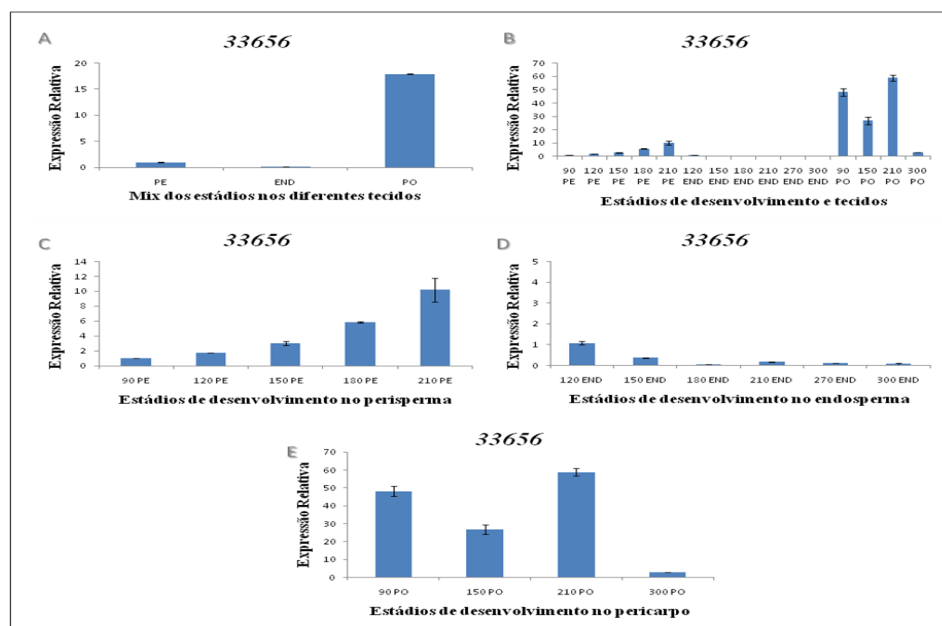


Figura 4 Expressão do *no hit* 33656 em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit* 33656 no mix dos estádios de desenvolvimento nos diferentes tecidos: perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit* 33656 nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit* 33656 nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit* 33656 nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit* 33656 nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador interno, em (B), (C), (D) e (E) o calibrador interno utilizado foi o 90PE.

A expressão do gene 33656 nos tecidos PE e END nas análises *in vivo* aqui apresentadas mostrou que este gene está mais expresso no estágio intermediário de desenvolvimento do fruto (210DAF) no perisperma, cerca de 10x mais do que a expressão no estágio inicial de desenvolvimento no

endosperma (120DAF), não corroborando com os dados *in silico* mostrados (Capítulo 2, Figura 5).

3.8 Expressão do gene *07128* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos

Os resultados das análises de qPCR mostraram que o gene *07128* parece ter uma expressão maior no endosperma (END) quando comparado ao perisperma (PE) e pericarpo (PO) (Figura 5). É possível observar que o gene *07128* apresentou uma expressão maior no endosperma quando comparado aos demais tecidos. Com relação à expressão deste gene no PE, é possível observar um aumento gradual da expressão ao longo do desenvolvimento do fruto até o estágio intermediário 180DAF. No entanto, aos 210DAF, a expressão deste *no hit* parece diminuir (Figura 5C). No endosperma (END), a expressão dele aumenta nos estádios intermediários (180DAF e 210DAF) e parece diminuir bruscamente nos estádios tardios (270DAF e 300DAF) (Figura 5D). No pericarpo (PO), parece haver um pico de expressão ao final do estágio inicial de desenvolvimento, 150DAF (Figura 5E).

As análises de qPCR mostraram que a expressão do gene *07128* foi maior no estágio intermediário de desenvolvimento (180DAF e 210DAF), corroborando com os resultados *in silico* apresentados (Capítulo 2, Figura 5).

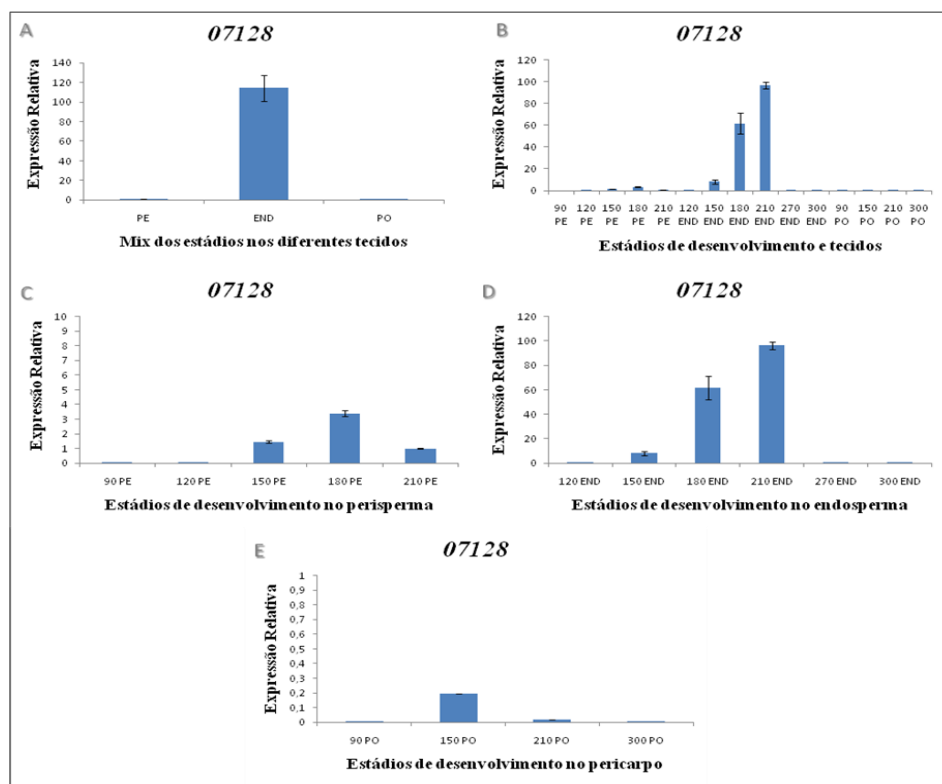


Figura 5 Expressão do *no hit 07128* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 07128* no mix dos estádios de desenvolvimento nos diferentes tecidos: perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 07128* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador, em (B), (C), (D) e (E) o calibrador utilizado foi o 210PE.

3.9 Expressão do gene *18112* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos

O resultados das análises de qPCR mostraram que o *no hit 18112* apresentou uma expressão maior no endosperma quando comparado aos demais tecidos (Figura 6). Com relação à expressão deste gene no perisperma, parece não haver oscilação da expressão deste gene ao longo do desenvolvimento do fruto (Figura 6C). No endosperma, é possível observar um aumento gradual da expressão ao longo do desenvolvimento do fruto (Figura 6D). No pericarpo, a expressão do *no hit 18112* também parece aumentar com o decorrer do desenvolvimento do fruto de *C. canephora* (Figura 6E).

A expressão do gene *18112* nas análises *in vivo* aqui apresentadas mostrou que este gene está mais expresso no estágio final de desenvolvimento do fruto (300DAF) tanto no endosperma quanto no pericarpo, contrastando com os resultados *in silico* apresentados (Capítulo 2, Figura 5).

Um perfil de expressão similar foi observado para os estádios finais de desenvolvimento do fruto de *C. arabica* para os genes pertencentes ao cluster I2, sendo eles: a dehidrin DH3, um transportador de hexoses (“*Hexose carrier protein*”), uma chitinase, uma forma 2 da sacarose sintase (“*Sucrose synthase sus2*”) e uma enzima da via de síntese dos ácidos clorogênicos (“*Cinnamoyl CoA reductase*”) (SALMONA et al., 2008).

É possível observar que o *no hit 18112* parece ser específico para o endosperma, ressaltando o interesse de analisar a tecido-especificidade do seu promotor que poderia ser usado como uma ferramenta para novos processos biotecnológicos (KOCHKO et al., 2010).

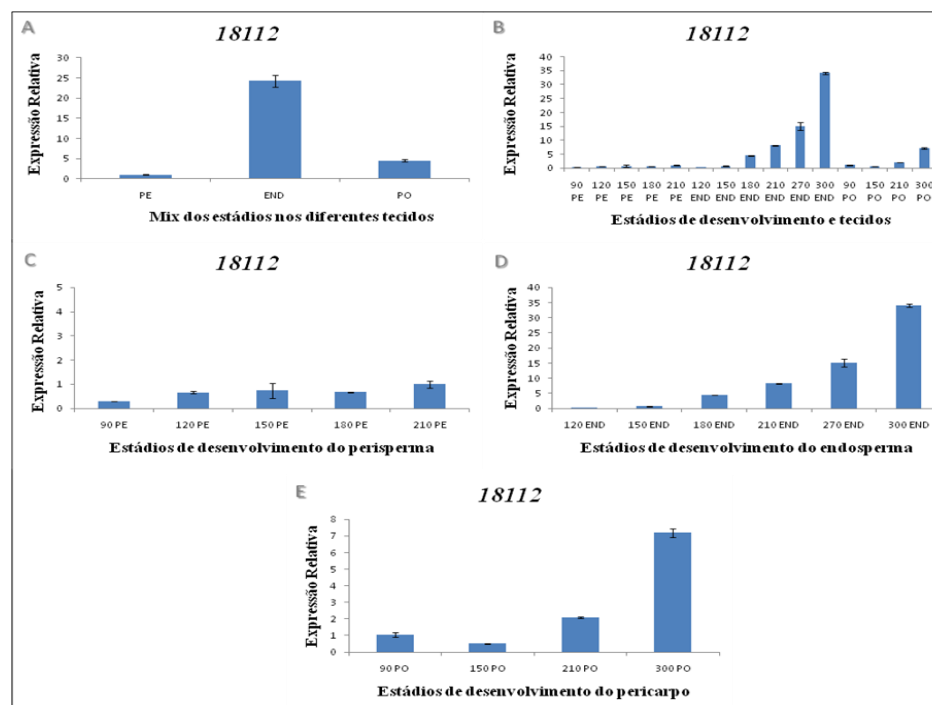


Figura 6 Expressão do *no hit 18112* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 18112* no mix dos estádios de desenvolvimento nos diferentes tecidos: perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 18112* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador interno, em (B), (C), (D) e (E) o calibrador interno utilizado foi o 210PE.

3.10 Expressão do gene *02344* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos

O *no hit 02344* apresentou uma expressão maior no perisperma (PE) quando comparado aos demais tecidos (Figura 7). Com relação à expressão deste gene somente no perisperma, é possível notar um aumento gradual da expressão gênica ao longo do desenvolvimento e maturação do fruto (Figura 7C). No endosperma (END), o *no hit 02344* parece ser mais expresso nos estádios 180DAF (intermediário) e 270DAF (tardio) (Figura 7D). No pericarpo (PO), a expressão desse *no hit* também parece aumentar no decorrer do desenvolvimento do fruto de *C. canephora*, porém diminui no estágio 300DAF (Figura 7E).

A expressão deste *no hit* para as análises de qPCR mostrou que este gene está mais expresso no estágio intermediário de desenvolvimento (180DAF e 210DAF) no perisperma, diminuindo consideravelmente a sua expressão nos estádios tardios de desenvolvimento do fruto (270DAF e 300DAF) no endosperma e no pericarpo, o que contrasta com as análise *in silico* apresentadas (Capítulo 2, Figura 5).

Como observado previamente para o *no hit 18112*, poderia ser de interesse analisar o promotor deste gene no âmbito de usá-lo como uma sequência promotora capaz de controlar a expressão de genes no perisperma de plantas transgênicas de café, uma vez que o *no hit 02344* parece ser específico para este tecido. Vale ressaltar, que nos estádios mais avançados do desenvolvimento do grão de café, (como a 210DAF), o perisperma está presente na forma de uma película extremamente fina composta de poucas células de espessura (GEROMEL et al., 2006). Os resultados obtidos aqui, mostrando um nível alto de expressão para *no hit* nestas poucas células do perisperma, reforçam o papel do perisperma na formação do grão de café, podendo este gene atuar neste processo.

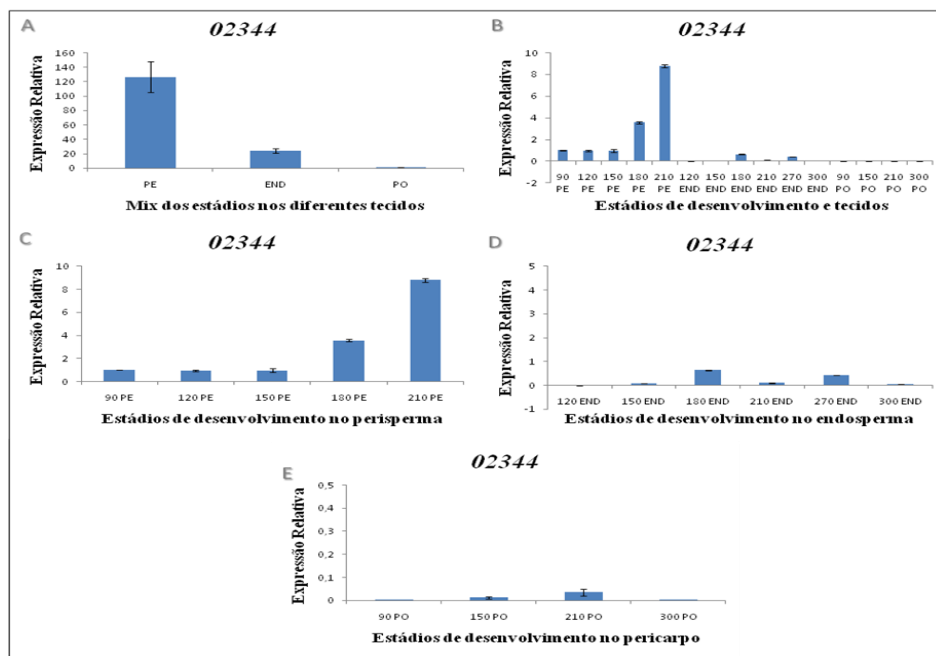


Figura 7 Expressão do *no hit 02344* em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit 02344* no mix dos estádios de desenvolvimento nos diferentes tecidos: perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit 02344* nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador interno, em (B), (C), (D) e (E) o calibrador interno utilizado foi o 90PE.

3.11 Expressão do gene 24645 em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos

É possível observar que o gene 24645 apresentou uma maior expressão no endosperma quando comparado aos demais tecidos (Figura 8). Com relação à expressão deste gene no perisperma, um aumento gradual da expressão até o estágio 150DAF é observado, onde este tecido volta a decrescer nos estádios 180DAF e 210DAF (Figura 8C). Entretanto, no endosperma, a expressão deste *no hit* é maior nos estádios 120DAF (inicial) e 300DAF (tardio) (Figura 8D). No pericarpo, parece haver um pico da expressão no estágio tardio 300DAF (Figura 8E).

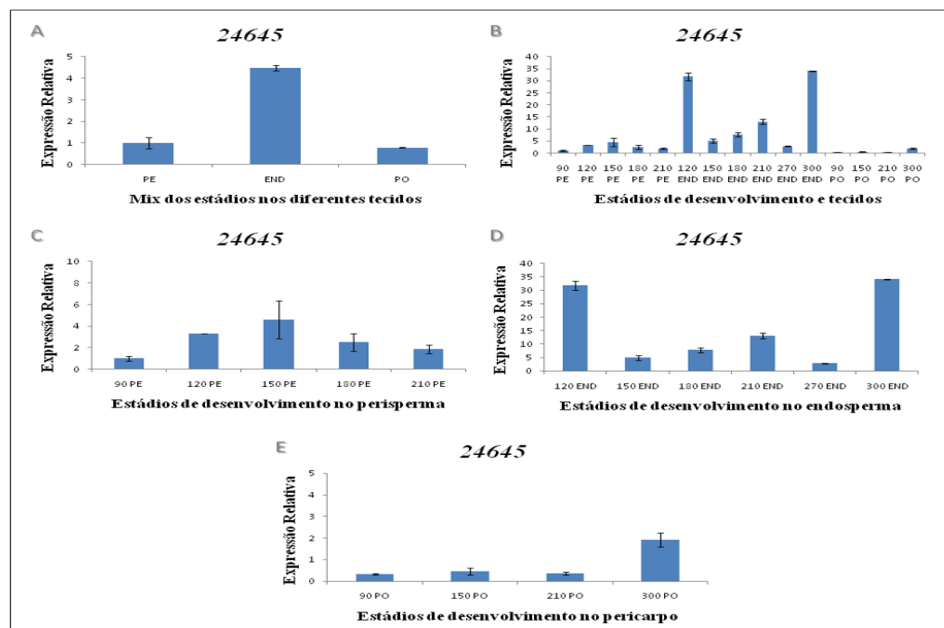


Figura 8 Expressão do *no hit* 24645 em frutos de *C. canephora* em diferentes estádios de desenvolvimento e tecidos. (A) Expressão relativa do *no hit* 24645 no mix dos estádios de desenvolvimento nos diferentes tecidos perisperma - PE, endosperma - END e pericarpo - PO. (B) Expressão relativa do *no hit* 24645 nos estádios de desenvolvimento inicial (90DAF-150DAF), intermediário (180DAF-210DAF) e tardio (270DAF-300DAF) dos diferentes tecidos PE, END e PO. (C) Expressão relativa do *no hit* 24645 nos estádios de desenvolvimento no perisperma. (D) Expressão relativa do *no hit* 24645 nos estádios de desenvolvimento no endosperma. (E) Expressão relativa do *no hit* 24645 nos estádios de desenvolvimento no pericarpo. O gene *CcUBQ10* foi utilizado como controle endógeno. Em (A) foi utilizado o PE como calibrador, em (B), (C), (D) e (E) o calibrador utilizado foi o 90PE.

3.12 Expressão dos genes *no hits* potencialmente envolvidos nos mecanismos de resposta à seca nas raízes de café

Os *no hits* 14149, 05517, 34630, 14549 e 42816 foram selecionados para analisar a expressão destes nas raízes dos clones de *C. canephora* tolerantes à seca 14, 73 e 120 e do clone 22 sensível à seca cultivados com (I) ou sem (NI) irrigação. Para esses *no hits*, pode-se notar que os resultados de expressão *in vivo* são altamente similares aos obtidos a partir das análises *in silico* (Figura 9). É possível observar que para todos os genes testados, o clone 73 não apresentou diferença de expressão significativa entre as duas condições analisadas (I e NI).

Já foram relatadas diferenças fisiológicas importantes entre os quatro clones analisados (MARRACCINI et al., 2011). Por exemplo, em relação à medida dos potenciais hídricos de antemanhã Ψ_{am} , foi observado que os clones atingiam o Ψ_{am} de -3,0 MPa em tempos diferentes um do outro. O clone 22 foi o primeiro a atingir o Ψ_{am} de -3,0 MPa em torno de 6 dias depois da suspensão de água, seguido dos clones 14 e 120 que atingiram este potencial juntos depois de aproximadamente 11 dias de seca, sendo que o clone 73 foi o último a atingir o estresse de -3,0 MPa 16 dias depois da suspensão de água.

3.13 Expressão do *no hit* 14149 em resposta à seca nas raízes de *C. canephora*

O *no hit* 14149 apresentou perfis de expressão diferencial entre os clones 14, 22 e 120 em resposta as condições de irrigação e não irrigação das plantas (Figura 9A). Assim para os clones 14, 22 e 120, foi possível observar um aumento da expressão deste *no hit* com o déficit hídrico aplicado para as plantas. Esta situação é mais evidente no clone 120 quando comparado aos demais. Para

o clone 73, não parece ter diferença de expressão significativa entre a condição de estresse com a seca (NI) ou sem estresse (I).

3.14 Expressão do *no hit 05517* em resposta à seca nas raízes de *C. canephora*

É possível observar perfis de expressão diferencial para o *no hit 05517* entre os clones 14, 22, 73 e 120 (Figura 9C), e também entre as condições de seca. Assim, a expressão deste *no hit* é claramente induzida com a seca nos clones 14, 22 e 120. No clone 73, a expressão deste *no hit* não apresenta variações significativas entre as duas condições de irrigação. O clone 120 apresentou uma expressão deste *no hit* na condição de seca, cerca de 170x maior do que a expressão nas plantas irrigadas deste mesmo clone. Os clones 22 e 14 apresentaram uma diferença relevante, também, quanto ao aumento da expressão na condição de estresse.

Neste ponto de vista, o perfil de expressão deste *no hit* é semelhante ao perfil de expressão do *no hit CcUNK10* observado nas folhas de *C. canephora* (MARRACCINI et al., 2012).

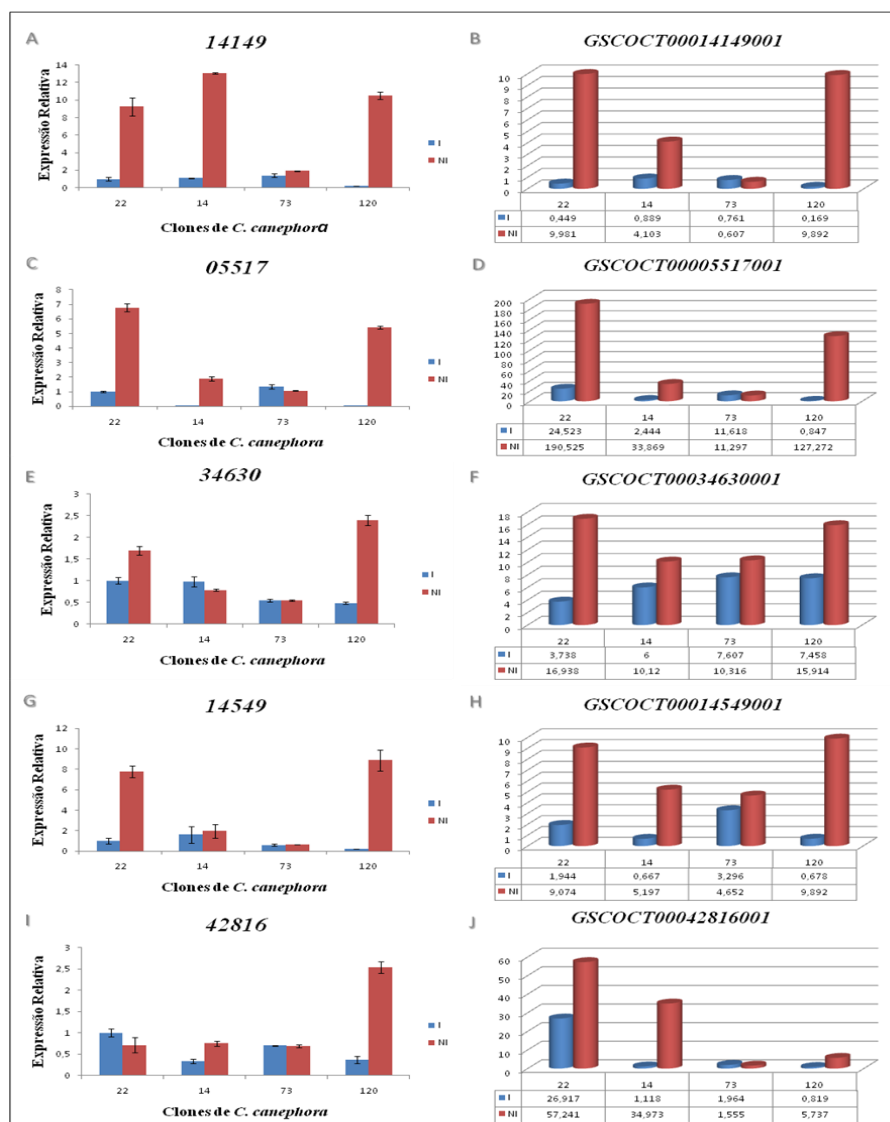


Figura 9 Perfis de expressão dos *no hits* 14149 (A, B), 05517 (C, D), 34630 (E, F), 14549 (G, H) e 42816 (I, J) na qPCR e *in silico*, respectivamente, nas raízes dos clones de *C. canephora* 14, 73 e 120 (tolerantes à seca) e o clone 22 (sensível à seca) submetidos ao estresse hídrico (NI: não irrigado) e sem estresse (I: Irrigado). Para as reações de qPCR o controle endógeno utilizado foi o *CcUBQ10*. O calibrador utilizado para normalização dos dados da qPCR foi o 22I. Nas análises *in silico* foi utilizado o BLASTp contra o banco de dados NR do NCBI, filtrados com E-value $\leq 1e^{-1}$.

3.15 Expressão do *no hit 34630* em resposta à seca nas raízes de *C. canephora*

Este *no hit* apresentou um aumento da expressão com a falta de água nos clones 22 e 120, sendo que o clone 120 apresentou uma diferença maior de expressão na condição de estresse, cerca de 5x mais, quando comparado à condição sem estresse do mesmo clone (Figura 9E). A expressão do *no hit 34630* nos clones 14 e 73 parece não oscilar quanto à presença do estresse de seca.

3.16 Expressão do *no hit 14549* em resposta à seca nas raízes de *C. canephora*

Para os clones 22 e 120, foi observada uma nítida expressão diferencial desse *no hit* com o estresse (Figura 9G). Assim, é possível observar que na condição de seca a expressão deste gene chegou a cerca de 24x mais no clone 120 quando comparado a condição sem estresse do mesmo clone. Já no clone 22, esta diferença chega a ser 5x maior na condição de estresse. Para os demais clones, 14 e 73, não foram observadas variações significativas da expressão do *no hit 14549* com as condições de irrigação e não irrigação. No entanto, é possível observar uma indução da expressão com o estresse hídrico do *no hit 14549* no clone 14 nas análises *in silico* apresentadas (Figura 9H).

3.17 Expressão do *no hit 42816* em resposta à seca nas raízes de *C. canephora*

Este *no hit* mostrou uma indução de expressão com o estresse hídrico nos clones 14 e 120 (Figura 9I). Essa indução foi maior (5x) nas raízes do clone

120 do que naquelas do clone 14. No clone 73, este gene não apresentou oscilação na expressão entre as duas condições analisadas. Já no clone 22 houve uma pequena diminuição da expressão do *no hit 42816* na condição de estresse hídrico. No entanto, é possível observar um aumento significativo da expressão com o estresse hídrico do *no hit 42816* no clone 14 nas análises *in silico* apresentadas, já o clone 120 não apresenta considerável oscilação da expressão deste *no hit* nas condições I e NI, contrastando com os resultados de qPCR (Figura 9J).

3.18 Expressão dos genes *no hits* potencialmente envolvidos nos mecanismos de resposta à seca no meristema de café

Os *no hits* selecionados para as análises com o estresse de seca nos meristemas de *C. arabica* foram 11524, 17768, 25639, 30926 e o 32791 (Figura 10), sendo que, com os resultados *in vivo* aqui apresentados, somente o *no hit 25639* esteve de acordo com as análises *in silico* (ver na Figura 7, Capítulo 2). Isto pode ser devido aos diferentes materiais utilizados para as análises *in silico* e posteriormente para as análises de qPCR. O material vegetal utilizado nas análises *in silico* foram coletados em setembro de 2009 onde os potenciais hídricos estavam em torno de -0,59 MPa para I59 e -1,20 MPa para Rubi, entretanto, para o experimento de qPCR mostrado neste trabalho, os meristemas foram coletados das cultivares em setembro de 2011 onde os potenciais das duas cultivares atingiram a -3,0 MPa.

3.19 Expressão dos *no hits* 11524 e 17768 em resposta à seca nos meristemas de *C. canephora*

Para esses dois *no hits*, foi observada uma queda do nível de expressão com a seca (Figuras 10A e B). Com irrigação, nota-se também que o *no hit* 11524 é mais expresso na cultivar I59 que em Rubi (Figura 10A). Uma situação oposta se observa na condição de não irrigação para o *no hit* 17768, que apresenta uma maior expressão em Rubi (cerca de 2x), do que a cultivar I59 nas mesmas condições (Figura 10B).

3.20 Expressão dos *no hits* 25639 e 30926 em resposta à seca nos meristemas de *C. canephora*

Ao contrario dos dois *no hits* precedentes, os *no hits* 25639 e 30926 apresentaram um aumento de expressão nos meristemas para as duas cultivares com o estresse hídrico (Figura 10C e 10D). Nas condições de seca, o nível da expressão do 25639 parece ser maior na cultivar Rubi que em I59 (Figura 10C) enquanto uma situação oposta se observa para o *no hit* 30926. Em relação aos níveis de expressão observados sem irrigação, pode-se notar que o aumento de expressão do 30926 parece ser maior (cerca de 6 até 10x, respectivamente em Rubi e I59) que o observado para o *no hit* 25639.

3.21 Expressão do *no hit* 32791 em resposta à seca nos meristemas de *C. canephora*

A expressão deste gene parece aumentar na cultivar Rubi submetida ao estresse de seca, entretanto para a cultivar I59 os níveis de expressão do gene

32791 parece não oscilar quanto as diferentes condições analisadas de irrigação e sem irrigação.(Figura 10D).

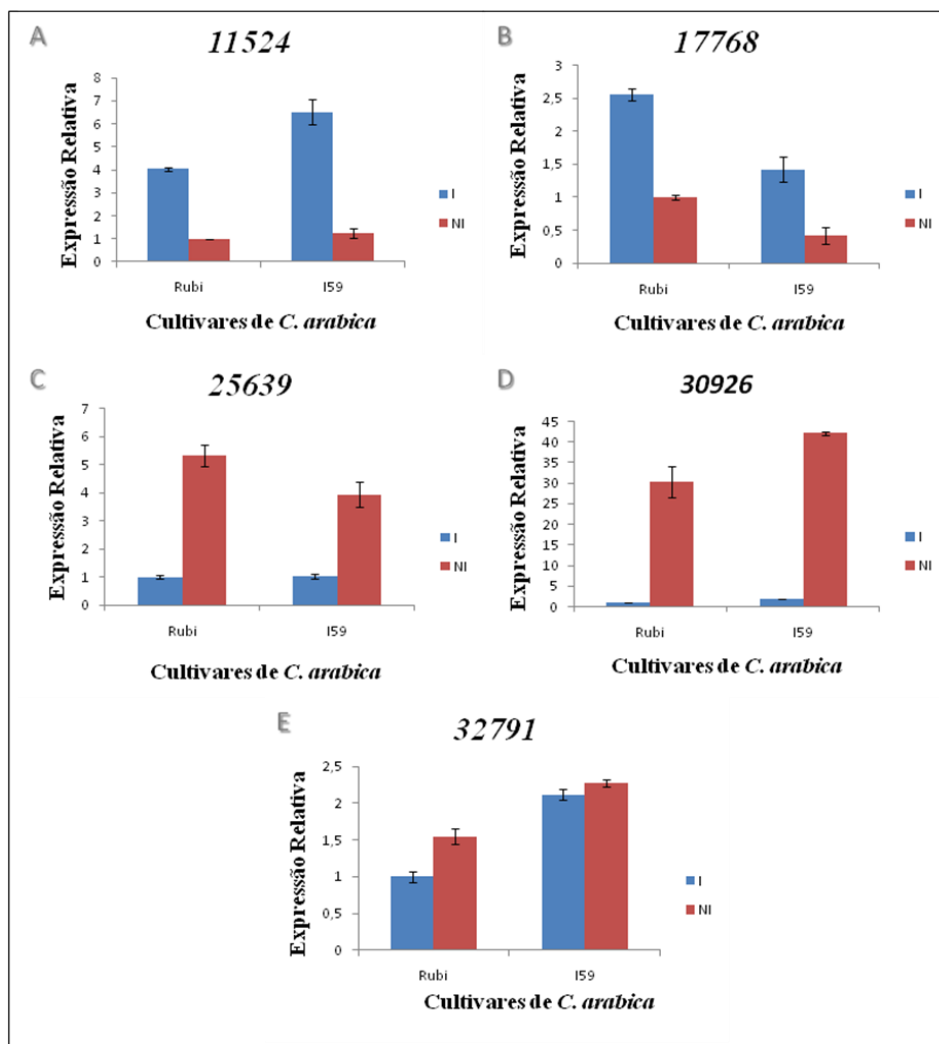


Figura 10 Perfis de expressão dos genes candidatos 11524 (A), 17768 (B), 25639 (C), 32791 (D) e 30926 (E) nos meristemas das cultivares Rubi e I59 de *C. arabica* submetidos ao estresse de seca (I: Irrigado) e sem estresse (NI: não irrigado). Rubi NI como calibrador em A e B e Rubi I como calibrador em C, D e E. O gene *CcUBQ10* foi utilizado como controle endógeno.

4 CONCLUSÃO

Na grande maioria dos *no hits* testados, foi possível validar a expressão dos genes selecionados a partir das análises *in silico*, por meio das análises de PCR quantitativo em tempo real, *in vivo*, com diferentes tecidos e condições de estresse aplicadas aos materiais vegetais de *C. arabica* e *C. canephora*. Assim, esse trabalho confirma a expressão (transcrição) de vários *no hits* de café. Foi possível caracterizar *no hits* potencialmente envolvidos na resposta da planta a estresses bióticos e abióticos, assim como *no hits* potencialmente envolvidos no desenvolvimento do fruto de café. Foi possível caracterizar *no hits* altamente específicos para determinado tecido no fruto de *C. canephora*, podendo indicar possíveis promotores para novos processos biotecnológicos com tecido-especificidade. 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 posteriores.

REFERÊNCIAS

ALBUQUERQUE, E. V. S. et al. Resistance to *Meloidogyne incognita* expresses a hypersensitive-like response in *Coffea arabica*. **European Journal of Plant Pathology**, Dordrecht, v. 127, n. 3, p. 365-373, 2010.

BARSALOBRES-CAVALLARI, C. F. et al. Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. **BMC Molecular Biology**, Berlin, v. 10, n. 1, p. 1-11, Jan. 2009.

BONETI, J. I. S.; FERRAZ, S. Modificação do método de Hussey & Barker para extração de ovos de *Meloidogyne incognita* de raízes de cafeeiros. **Fitopatologia Brasileira**, Brasília, v. 6, p. 553, 1981.

CARNEIRO, R. M. D. G. et al. Reação de cafeeiros conilon a diferentes populações de *meloidogyne* spp. In: SIMPÓSIO DOS CAFÉS DO BRASIL, 6., 2009, Vitória. **Anais...** Brasília: Consórcio Pesquisa-Café, 2009. 1 CD-ROM.

FLEGG, J. J. M. Extraction of *Xiphinema* and *Longidorus* species from soil by a modification of Cobb's decanting and sieving technique. **Annals of Applied Biology**, Warwick, v. 60, p. 429-437, 1967.

GEROMEL, C. et al. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. **Journal of Experimental Botany**, Oxford, v. 57, n. 12, p. 3243-3258, Sept. 2006.

HUSSEY, R. S.; BARKER, K. R. A comparison of methods of collecting inocula of *Meloidogyne* spp. including a new technique. **Plant Disease Reporter**, Washington, v. 57, p. 1025-1028, 1973.

KOCHKO, A. de et al. Advances in *Coffea* genomics. In: KADER, J. C.; DELSENY, M. (Ed.). **Advances in botanical research**. Oxford: Academic, 2010. p. 23-53.

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**, Berlin, 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**, Oxford, v. 63, n. 11, p. 4191-4212, 2012.

_____. RBCS1 expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. **BMC Plant Biology**, Berlin, v. 11, n. 85, p. 1-23, May 2011.

PINHEIRO, H. A. et al. Drought tolerance is associated with rooting depth and stomatal control of water use in clones of *Coffea canephora*. **Annals of Botany**, London, v. 96, n. 1, p. 101-108, July 2005.

SALMONA, J. et al. Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. **Plant Molecular Biology**, Berlin, v. 66, n. 1/2, p. 105-124, 2008.

VIDAL, R. et al. High-throughput sequencing of cDNA shows that cv. Rubi and IAPAR59 of *Coffea arabica* have different molecular response to water privation. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 24., 2012, San José. **Proceedings...** San José: ICCS, 2012. 1 CD-ROM.