



**LUCIANA PEREIRA FREIRE**

**ESTUDO DA EXPRESSÃO E ANÁLISE DE  
POLIMORFISMOS DE GENES CANDIDATOS  
PARA A TOLERÂNCIA À SECA EM CAFEIEIRO**

**LAVRAS - MG**

**2010**

**LUCIANA PEREIRA FREIRE**

**ESTUDO DA EXPRESSÃO E ANÁLISE DE POLIMORFISMOS DE  
GENES CANDIDATOS PARA A TOLERÂNCIA À SECA EM  
CAFEIEIRO**

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

Orientador

Dr. Pierre Roger René Marraccini

Coorientador

Dr. Alan Carvalho Andrade

**LAVRAS-MG**

**2010**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Freire, Luciana Pereira.

Estudo da expressão e análise de polimorfismos de genes  
candidatos para a tolerância à seca em cafeeiro / Luciana Pereira  
Freire. – Lavras : UFLA, 2010.

117 p.: il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2010.

Orientador: Pierre Roger René Marraccini.

Bibliografia.

1. Coffea. 2. Expressão gênica. 3. Marcadores moleculares. 4.  
SNP. 5. Estresse abiótico. I. Universidade Federal de Lavras. II.  
Título.

CDD – 583.520487328

**LUCIANA PEREIRA FREIRE**

**ESTUDO DA EXPRESSÃO E ANÁLISE DE POLIMORFISMOS DE  
GENES CANDIDATOS PARA A TOLERÂNCIA À SECA EM  
CAFEIEIRO**

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

APROVADA em 31 de agosto de 2010.

Dr. Alan Carvalho Andrade      Embrapa Cenargen

M. Sc. Gustavo Costa Rodrigues      Embrapa Informática Agropecuária

Dr. Pierre Roger René Marraccini  
Orientador

Dr. Alan Carvalho Andrade  
Coorientador

**LAVRAS - MG**

**2010**

***Dedico***

À minha família, principalmente a minha mãezinha, Hercília e a minha irmã Cristina, pelo apoio, incentivo, paciência, enfim por estarem sempre ao meu lado, principalmente nas horas mais difíceis. E a todos meus amigos. Amo todos vocês!

## AGRADECIMENTOS

A Deus, por estar sempre ao meu lado iluminando meu caminho para que eu possa superar as adversidades.

A minha família, meus pais Antônio e Hercília, e meus irmãos pela confiança orações e amor. Sem vocês não seria possível concretizar esse trabalho.

Aos meus amigos e companheiros de laboratório por toda, amizade, apoio e sugestões e por fazerem do ambiente de pesquisa um ótimo lugar para se trabalhar: Mariana, Marry, Maria, Hugo, Bárbara, André, Gabi e em especial agradeço ao Lipe, ao Gabriel, ao Fabrício, a Nat, a Ingrid, a Betúlia e ao Eder, pelos momentos de descontração e encorajamento e por todo trabalho feito para que eu pudesse realizar o meu. Vocês são ótimos!

As minhas irmãs de república: Joyce, Dani e Cássia, pela grande amizade, companhia, e divertida convivência.

Aos amigos do mestrado e agora amigos para vida toda: Nadinha, Brenda, Glacy, Fabiana, Romário, Gustavo, Kátia, Rodrigo e Horllys. Obrigada pela torcida e por dividir comigo as mesmas angústias e principalmente pela amizade.

As minhas amigas Geiza e Aline, pela amizade verdadeira e por serem tão especiais.

A todos os meus amigos da Embrapa: Bruninha, Tiago Rafaela, Rafael, Lara. As meninas do laboratório de Apomixia: Larissa e Aninha, ao pessoal da plataforma de seqüenciamento: Mário e Lú que nunca mediram esforços para poder me ajudar.

A minha amiga Val que me acompanha desde a graduação, obrigada por toda convivência e paciência.

Ao meu mais novo amigo Martins por toda paciência, apoio e ajuda.

Aos professores do curso de Biotecnologia Vegetal, pelos valiosos ensinamentos.

Ao professor Luciano, coordenador do curso de Biotecnologia Vegetal, pela oportunidade.

Ao meu orientador Pierre, pela oportunidade de realizar esse trabalho, pela paciência e principalmente por todos os ensinamentos que são eternos! E que por mais ocupado que estivesse sempre estava disposto a me atender. Obrigada!

Ao meu co-orientador Alan, pela oportunidade de realizar esse trabalho, ensinamentos e confiança depositada ao longo destes anos de trabalho. Obrigada!

Aos membros da banca examinadora, Dr. Alan, Dr. Pierre e M. Sc Gustavo por aceitarem o convite para fazer parte da banca.

Enfim, a todos que, de alguma forma, contribuíram para a realização desse trabalho.

“ Agradecer é admitir que houve um momento em que se precisou de alguém.

É reconhecer que o homem jamais poderá ser auto-suficiente.

Ninguém cresce sozinho, sempre é preciso um olhar de apoio uma palavra de incentivo, um gesto de compreensão. Agradeço então a todos vocês que de alguma forma participaram dos meus ideais.”

(Autor desconhecido)

## RESUMO

O cafeeiro é uma planta perene, pertencente à família Rubiaceae e ao gênero *Coffea*. Entre as espécies cultivadas, *Coffea canephora* (Robusta) e *Coffea arabica* (Arabica) são as mais importantes economicamente, representando respectivamente 30% e 70% da produção comercial mundial. O déficit hídrico afeta a cultura do cafeeiro podendo resultar em importantes conseqüências sociais (deslocamento de trabalhadores), econômicas e ecológicas. Estudos na área de melhoramento do café têm visado à introdução de novos caracteres para a obtenção de híbridos mais tolerantes à seca. Em nível molecular, existem projetos para se identificar genes candidatos (GC) para a tolerância à seca, seja por meio de estudos da expressão gênica (transcriptoma) ou de proteínas (proteômica). Assim, o primeiro objetivo desse trabalho consistiu em estudar a expressão de GCs (*DREB*, *NAC-R26*, *RD29*, *GC10*, *CCoAMT*, *OEC*, *CA*, *M6FR*, as isoformas *RBCS1-A* e *B* do gene *RBCS*), utilizando-se vários cultivares de *C. arabica* cultivados em campo sob situações diversas de estresse hídrico, por meio da técnica de PCR em tempo real (qPCR). Os GCs utilizados foram previamente identificados e validados, em experimentos com cultivares de *C. canephora* cultivados em casa de vegetação. O segundo objetivo consistiu em analisar a diversidade nucleotídica *in vivo* de alguns GCs (*DREB*, *RD29* e *NAC-RD26*), identificando-se SNPs (*Single Nucleotide Polymorphisms*), em várias espécies, clones e cultivares de café para tentar relacionar a ocorrência de SNPs com o fenótipo. Para a busca de SNPs, DNA genômico (gDNA) dessas diferentes plantas foram amplificados com dois *primers* específicos para cada GC utilizado. Os primeiros *primers* contêm nas extremidades adaptadores M13For ou M13Rev, os quais permitiram avaliar a presença de SNPs, por meio do seqüenciamento dos produtos de PCR sem a clonagem. Para se identificar as formas alélicas presentes em cada genoma, *primers* (sem adaptadores) foram usados para se amplificar, clonar e seqüenciar as mesmas porções de gDNA. Para o seqüenciamento, foram utilizados 10 clones independentes, visando-se a identificação do máximo possível de alelos. As seqüências resultantes foram alinhadas, pelo emprego de aplicativos computacionais visando-se a identificação de regiões polimórficas.

Palavras-chave: *Coffea*. Expressão gênica. Marcadores moleculares. SNP. Estresse abiótico.

## ABSTRACT

The coffee is a perennial tree which belongs to the Rubiaceae family and the genus *Coffea*. Among the cultivated species, *Coffea canephora* (Robusta) and *Coffea arabica* (Arabica) are the most important economically, representing respectively 30% and 70% of the world trade production. One major problem that affects this culture is its susceptibility to soil water deficiency, which can lead to important economic, ecological and social consequences – such as the massive migration of unemployed rural workers to urban centers. One of the objectives of coffee genetic improvement is to produce drought-tolerant coffee hybrids. At the molecular level, some projects are being carried out in order to identify candidate genes (CG) for drought tolerance, either by the evaluation of differential gene expression on both the transcriptional (transcriptome) and the translational levels (proteomics). The first objective of this work consisted in the evaluation by the real-time PCR (qPCR) technique, of the differential expression of drought-tolerance CGs, such as *DREB*, *NAC-R26*, *RD29*, *GC10*, *CCoAMT*, *OEC*, *A.C*, *M6FR*, as well as the A and B isoforms of the *RBCS* gene, in several *C. arabica* cultivars grown under different field conditions of drought intensity. The CGs were previously identified and validated in experiments with cultivars of *C. canephora* grown in a greenhouse. The second objective was to analyze the *in vivo* nucleotide diversity of some CGs (*DREB*, *RD29* and *RD26-NAC*) present in several varieties of coffee, and to identify SNPs (Single Nucleotide Polymorphisms) related to the drought tolerance phenotype. The genomic DNAs (gDNAs) from different species, clones and cultivars of coffee were extracted and submitted to specific amplification with two CGs primers carrying forward and reverse M13 adapters in the extremities, which allowed us to evaluate the presence of SNPs by sequencing the PCR products without cloning. To identify the allelic forms present in each genome, regular primers were utilized to amplify and to sequence the same portions of gDNA after proper cloning. The resulting allelic sequences of ten independent clones were aligned *in silico* and different polymorphic regions were properly identified.

Keywords: *Coffea*. Gene expression. Molecular markers. SNP. Abiotic stress.

## SUMÁRIO

1	INTRODUÇÃO.....	12
2	REFERENCIAL TEÓRICO .....	18
2.1	A cultura do café .....	18
2.2	Importância econômica do café .....	20
3	ESTRESSE HÍDRICO (ASPECTOS GERAIS) .....	22
3.1	Respostas ao estresse abiótico em plantas .....	23
3.2	Respostas ao estresse hídrico em <i>Coffea</i> .....	28
3.3	Café e biotecnologia .....	29
3.4	Marcadores moleculares .....	31
3.5	SNPs (Single Nucleotide Polymorphisms, INDELS – Insertion/Deletions).....	32
4	MATERIAL E MÉTODO.....	35
4.1	Validação da expressão de genes candidatos em cultivares de <i>C.</i> <i>arabica</i> cultivadas no campo .....	35
4.1.1	Caracterização fisiológica .....	36
4.1.2	Extração de RNA .....	37
4.1.3	Análise da qualidade e quantificação das alíquotas de RNA .....	38
4.1.4	Tratamento com DNase .....	38
4.1.5	Transcriptase reversa (RT).....	39
4.1.6	Desenho de primers para as análises de expressão pela qPCR.....	40
4.1.7	Análises de expressão pela PCR quantitativo em Tempo Real (qPCR) .....	41
4.1.7.1	Avaliação da eficiência dos <i>primers</i> .....	42
4.1.7.2	Análises da expressão por PCR quantitativo em tempo real .....	42
4.2	Busca de SNPs .....	43
4.2.1	Desenho de <i>primers</i> para amplificação dos genes utilizados na busca de SNPs. ....	45
4.2.2	Reação em cadeia da polimerase (PCR) e amplificação do gDNA .	46
4.2.3	Purificação de fragmentos de gDNA a partir do gel de agarose.....	48
4.2.4	Ligação dos fragmentos purificados e clonagem.....	49
4.2.5	Transfecção bacteriana. ....	51
4.2.6	Extração de DNA plasmidial “Miniprep” .....	51
4.2.7	Sequenciamento .....	52
4.2.8	Análise das sequências.....	54
5	RESULTADOS E DISCUSSÃO.....	55
5.1	Caracterização fisiológica .....	55
5.2.1	Expressão do gene <i>NAC-RD26</i> .....	59
5.2.2	Expressão do gene <i>RD29</i> .....	59
5.2.3	Expressão do gene Mannose-6-fosfato redutase ( <i>M6FR</i> ) .....	60

5.2.4	Expressão do gene <i>GCI0</i> .....	60
5.2.5	Expressão dos genes <i>AC</i> e <i>OEC</i> .....	62
5.2.6	Expressão do gene <i>DREB</i> .....	62
5.2.7	Expressão do gene <i>CCoAMT</i> .....	62
5.2.8	Expressão do gene <i>MYB61</i> .....	65
5.2.9	Expressão do gene <i>RBCS</i> .....	65
5.2.10	<b>Discussão: validação da expressão dos genes candidatos em cultivares de <i>C. arabica</i> cultivados no campo</b> .....	67
5.2.10.1	Efeito do estresse hídrico nos dois anos de estudo sobre a expressão dos GC.....	67
5.2.10.2	Comparação dos perfis de expressão dos GC entre os cultivares <b>I59</b> e <b>Rubi</b> de <i>C. arábica</i> .....	68
5.2.10.3	Comparação dos perfis de expressão dos GC em <i>C. arabica</i> e em <i>C. canephora</i> .....	69
5.2.10.4	Caso particular do gene <i>RBCS</i> .....	71
5.3	<b>Análise dos polimorfismos (SNPs) do gene <i>DREB</i></b> .....	72
5.3.1	<b>Análise dos polimorfismos (SNPs) do gene <i>NAC-RD26</i></b> .....	77
5.3.2	<b>Análise dos polimorfismos (SNPs) do gene <i>RD29</i></b> .....	82
5.3.3	<b>Discussão das análises dos polimorfismos (SNPs)</b> .....	89
6	<b>CONCLUSÕES E PERSPECTIVAS</b> .....	95
	<b>REFERÊNCIAS</b> .....	98
	<b>APÊNDICE</b> .....	109

## 1 INTRODUÇÃO

Nos próximos anos, a produção de alimentos pode diminuir no mundo dando lugar às culturas que estão voltadas para a produção de agroenergia, pois culturas mais ligadas à alimentação, como o café, a soja e o milho são mais vulneráveis aos efeitos do aquecimento global. Países pobres da África e da Ásia seriam os mais afetados, mas grandes produtores agrícolas, como o Brasil, também sentiriam os efeitos já na próxima década (ASSAD et al., 2001).

Se nada for feito para diminuir ou amenizar os efeitos das mudanças climáticas visando à adaptação das culturas para a nova situação, a geografia da produção agrícola no Brasil poderá mudar drasticamente nos próximos anos, mesmo com o aquecimento limitado em 1,5 °C até o final do século (ASSAD, 2009). De fato, isso trará importantes consequências sociais (deslocamento de trabalhadores), econômicas e ecológicas.

Assim, no Brasil o aquecimento global poderá resultar em perdas de começam até R\$ 7,4 bilhões em 2020. A soja de vera ser a cultura mais afetada, as perdas poderão chegar a 40% em 2070. A mandioca pode desaparecer do Semi-árido nordestino, o milho, o arroz, o feijão, o algodão e o girassol sofrerão forte redução de área no Nordeste. Nesse cenário, o café Arabica também terá poucas condições de sobrevivência no Sudeste (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA, 2009).

Vários autores já constataram os efeitos benéficos em termos de produtividade em cafeeiros irrigados quando comparados a não-irrigados (ANTUNES et al., 2000; SOARES et al., 2000). Afirmam que a irrigação propicia maior produtividade. A falta ou o excesso de água certamente afeta o desenvolvimento das plantas. Por exemplo, o cafeeiro, em consequência da falta de água no solo, tem seu metabolismo alterado. Há redução do fluxo de vapor e da transpiração, bem como da absorção de água e de nutrientes pelo sistema

radicular, especialmente pelas raízes absorventes e, conseqüentemente, diminuição da produção (MATIELLO; DANTAS, 1987).

Atualmente, estudos na área de melhoramento do café tem visado à introdução de novos caracteres para obtenção de híbridos mais tolerantes a estresses bióticos (doenças: ferrugem e pragas: bicho mineiro) e a estresses abióticos (seca sendo este o foco desse trabalho), mas se conhece relativamente pouco sobre como os genótipos de café respondem ao estresse hídrico nos níveis morfológicos, fisiológicos, e molecular (DAMATTA; RAMALHO, 2006).

Em nível molecular, existem projetos para identificar os genes candidatos (GC) em café, tais como os projetos em andamento no Laboratório de Genética Molecular (LGM) na Embrapa Recursos Genéticos e Biotecnológicos (ANDRADE et al., 2006; VINECKY, 2008), em parceria com o Instituto Agrônomo do Paraná (IAPAR), o Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) e o Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER). Alguns GC já foram identificados, seja através da expressão diferencial no nível da transcrição (transcriptoma) ou das proteínas (proteômica) (ALVES et al., 2009; RAMOS et al., 2008).

Genes candidatos (GC) são genes relacionados a uma dada característica agronômica de interesse, por exemplo, conferir uma tolerância ou resistência aos estresses bióticos e abióticos (VARSHHEY; GRANER; SORRELLS, 2004).

O Instituto Capixaba de Assistência Técnica e Extensão Rural (INCAPER) identificou várias plantas, de *C. canephora* var. Conilon tolerantes (como os clones 14, 73 e 120) e sensível (como o clone 22) ao déficit hídrico (FERRÃO et al., 1999). Mudanças desses clones foram submetidas ao estresse hídrico em casa de vegetação e caracterizadas fisiologicamente na Universidade Federal de Viçosa (UFV) (DAMATTA et al., 2003; LIMA, 2002; PINHEIRO, 2005; PRAXEDES et al., 2006). Posteriormente folhas e raízes foram utilizadas

em estudos para a caracterização do perfil da expressão gênica os clones 14 e 22, tolerante e sensível respectivamente, cultivados sob condições controle e de estresse hídrico (SILVA, 2007).

Assim, foi possível a identificação de vários GC já descritos na literatura como essenciais às repostas das plantas ao déficit hídrico:

- a) Gene *AC* que codifica para uma anidrase carbônica do cloroplasto necessária para a solubilização celular do  $\text{CO}_2$  (na forma de  $\text{HCO}_3^-$ ), substrato da RUBISCO;
- b) Gene *OEC* (“*Oxygen evolving enhancer protein*”) que codifica para uma das três proteínas do complexo de evolução de oxigênio do PSII dos cloroplastos;
- c) Gene *M6FR* codifica para a NADPH-manose dependente -6- fosfato redutase (EC 1.1.1.224), enzima chave na produção de manitol convertendo o manose-6-fosfato em manitol-1-fosfato. Plantas transgênicas que acumulam manitol parecem ser mais tolerantes ao estresse salino (ZHIFANG; LESCHER, 2003);
- d) O gene *GC10* codifica para uma proteína de função desconhecida. Esse gene é altamente expresso nas folhas do clone 22 (sensível a seca) de *C. canephora* em condição de estresse hídrico (MARRACCINI et al., 2008);
- e) Gene *RD29* (“responsive drought”) é induzido em todos os tecidos da planta, incluindo folhas e raízes, órgãos importantes na percepção e sinalização do estresse, durante períodos de déficit hídrico, entretanto a função precisa da proteína desse gene não é conhecida (KANG et al., 2002);
- f) Genes *DREB* (“Dehydration-responsive element binding protein”). Os genes *DREB* codificam para fatores de transcrição que reconhecem

especificamente os elementos regulatórios *DRE* nas regiões promotoras de genes expressos em condições de estresse hídrico em várias espécies, tais como *Arabidopsis*, trigo, centeio, milho arroz e cevada (LIU, 2006). Esses genes podem ser divididos em dois grupos de acordo com o estresse envolvido: *DREB1* relacionado com a tolerância à baixas temperaturas (frio) e *DREB2* associado à tolerância a seca;

- g) Genes *NAC* codificam para fatores de transcrição implicados no controle de diversos processos nas plantas tais como o desenvolvimento, defesa e as respostas aos estresses abióticos (OLSEN et al., 2005);
- h) Genes *MYB* codificam para fatores de transcrição, controlando desenvolvimento e as respostas das plantas aos fatores ambientais, à ação mediada por hormônios e à regulação do metabolismo dos fenil propanoides (TRAN et al., 2007);
- i) Gene *CCoAMT* codifica para a enzima caffeoyl-CoA O-methyltransferase (EC. 2.1.1.1 04) implicada na via de biossíntese dos ácidos clorogênicos e das ligninas (LI, 1999);
- j) Gene *RBCS* que codifica para a RUBISCO (ribulose-bisfosfato carboxilase oxigenase, EC 4.1.1.39). Essa enzima cloroplástica atua na fixação do CO<sub>2</sub>, adicionando um átomo de carbono no ribulose 1,5-difosfato. No núcleo, existe uma família de genes *RBCS*, cada um codificando para isoformas da RUBISCO que se expressam em condições específicas (MARRACCINI et al., 2003).

Dessa forma, ao observar os resultados obtidos para o clone tolerante (clone 14) e sensível (clone 22) à seca de *C. canephora* cultivados em casa de vegetação em diferentes regimes hídricos, decidimos trabalhar com esses GC, pois o primeiro objetivo desse trabalho é estender essas validações para vários

cultivares de *C. arabica* cultivados no campo. Desta forma foi possível comparar a expressão dos mesmos genes já estudados em cultivares de *C. canephora*, utilizando-se agora plantas de *C. arabica* sob condições diversas de estresse hídrico (Figura 1).

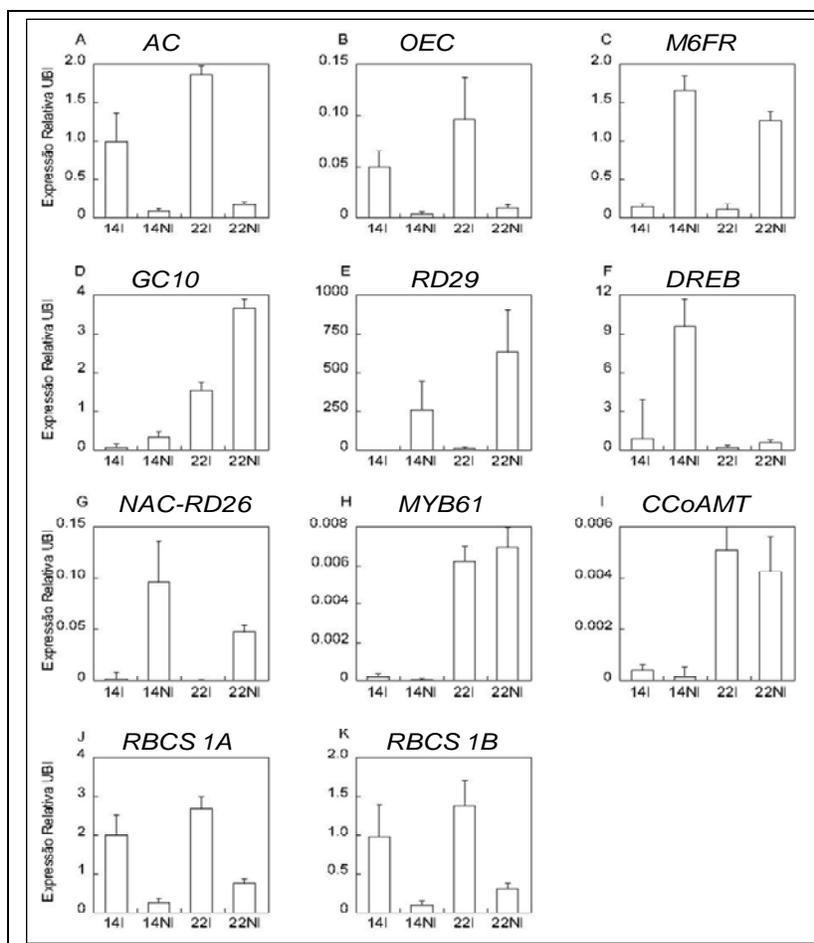


Figura 1 Perfis de expressão dos GC em folhas dos clones 14 e 22 de *C. canephora* var. Conilon com Irrigação (I) e sem Irrigação (NI), obtidos por meio da técnica de qPCR. Genes com aumento de expressão sob limitação de água *M6FR* (C), *GC10* (D), *RD29* (E), *DREB* (F), *NAC-RD26* (G), *MYB61* (H). Genes com redução de expressão sob limitação de água: *AC* (A) e *OEC* (B), *CCoAMT* (I), *RBCS1A* (J), *RBCS1B* (K)

Fonte: Marraccini et al. (2009)

O segundo objetivo desse trabalho foi analisar a diversidade nucleotídica *in vivo* de alguns GC por meio da identificação SNPs (*Single Nucleotide Polymorphisms*) em diferentes cultivares de cafeeiro. Como algumas plantas analisadas apresentam comportamentos diferentes em relação à tolerância a seca (tal como as plantas dos subgrupos SG1 e SG2 do grupo de congolês de *C. canephora*, tentamos relacionar as análises das sequências nucleicas e polimorfismos (SNPs) identificados com o fenótipo observado dessas plantas.

## 2 REFERENCIAL TEÓRICO

### 2.1 A cultura do café

O cafeeiro é uma planta perene, pertencente à família Rubiaceae e ao gênero *Coffea*. Entre as espécies cultivadas, *Coffea arabica* (Arabica) e *Coffea canephora* Pierre (Robusta) são as mais importantes economicamente, representando aproximadamente e respectivamente 70% e 30% da produção comercial mundial (INTERNATIONAL COFFEE ORGANIZATION - ICO, 2009). A primeira espécie é reconhecida por produzir grãos que dão bebida de boa qualidade ou café de coador, com baixos teores de cafeína. A segunda espécie é utilizada como matéria-prima para o café solúvel, produzindo grãos considerados de menor qualidade com maiores teores de cafeína e ácidos (LEROY et al., 2006).

*C. canephora* Pierre é uma espécie alógama, diploide, ( $2n=2x=22$  cromossomos). É originária de regiões quentes e úmidas, de baixa altitude, da costa oeste à região central do continente Africano, que se estende da Guiné ao Congo (MENDES; GUIMARÃES; SOUZA, 2002). Ao contrário, *C. arabica* L. ( $2n=4x=44$ ) é autógama e alotetraploide contendo dois subgenomas diploides diferentes. A origem mais provável de *C. arabica* é a hibridação natural entre as espécies diploides *C. eugenioides* e *C. canephora* (LASHERMES et al., 1999).

Apesar de somente *C. arabica* e *C. canephora* terem importância econômica no mercado mundial, outros acessos de espécies selvagens não cultivadas comercialmente, apresentam características altamente vantajosas, tais como resistência ao bicho mineiro no caso de *C. racemosa* e à ferrugem no caso do Híbrido de Timor. Programas de melhoramento genético do cafeeiro, além de visar o desenvolvimento de cultivares produtivas, como resistência à ferrugem e nematoides, focam também na obtenção de variedades de café tolerantes à seca

(FERRÃO et al., 1999). Algumas espécies tais como *C. canephora*, *C. racemosa* e *C. congensis* vêm sendo utilizadas como fontes importantes de variabilidade genética nestes programas (BERTRAND et al., 2000). Entretanto devido ao modo de reprodução, a diversidade genética presente na espécie *C. canephora* é mais ampla que na espécie de *C. arabica*. Análises da diversidade genética de plantas de *C. canephora* por meio de marcadores moleculares nos mostram a existências de dois grupos distintos (Figura 2), estabelecidos em função da sua origem geográfica denominados: Guineano e Congolês (MONTAGNON; LEROY; YAPO, 1992).

O grupo Guineano é constituído por populações selvagens da Costa do Marfim 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 freqüentemente bronze (DUSSERT et al., 1999; FAZUOLI, 2007; MONTAGNON; LEROY; YAPO, 1992).

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 a seca que corresponde ao café Robusta propriamente dito. 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 (DUSSERT et al., 1999; FAZUOLI, 2007; MONTAGNON; LEROY; YAPO, 1992).

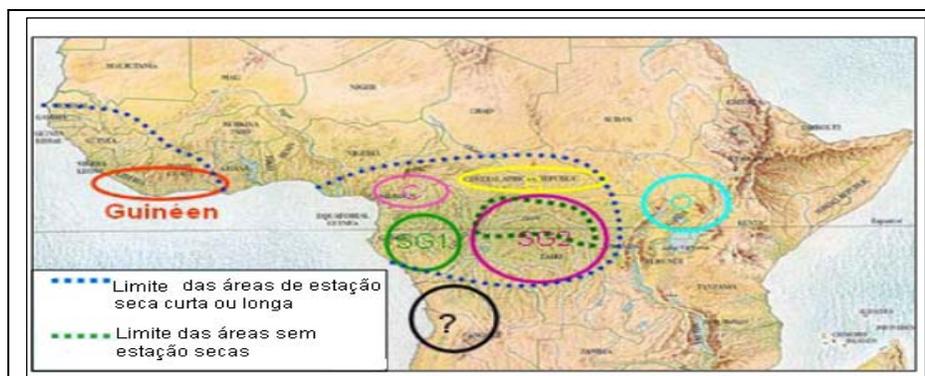


Figura 2 Grupos de distribuição da diversidade de *Coffea canephora*  
 Fonte: Montagnon (2000)

## 2.2 Importância econômica do café

Dentre as *commodities* naturais, o café tem seu valor ultrapassado apenas pelo petróleo, pois é a segunda mercadoria mais comercializada do mundo (DAMATTA; RAMALHO, 2006). No Brasil, tal como em outros países produtores do mundo o agronegócio do café é responsável pela geração de um grande número de empregos, (oito milhões de pessoas em todos os setores da economia), indo desde os setores de máquinas, equipamentos e insumos, passando pela produção no campo e pela indústria, até o setor de serviços, como logística e comércio.

Os maiores produtores mundiais de café na ordem decrescente são: o Brasil, o Vietnã, a Colômbia, a Indonésia, a Índia, o México, a Etiópia, a Guatemala, a Uganda e o Peru (ICO, 2009).

No Brasil, a área plantada está distribuída por todo território. O Brasil é líder mundial nas exportações do produto (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2009). Embora tenha ocorrido uma redução da área cultivada nas últimas décadas, devido a diversos fatores abióticos como seca e geada, de fatores bióticos como pragas e doenças (GONÇALVES et al.,

1998). Em relação ao mercado consumidor, o Brasil é o 2º mercado do mundo, atrás apenas dos EUA (ICO, 2009). De acordo com a CONAB (2009), o país possui aproximadamente 6,4 bilhões de pés de café, com 2,3 milhões de hectares de área plantada, pouco mais da metade só no Estado de Minas Gerais.

Em termo da produção nacional, o Estado de Minas Gerais possui as maiores plantações de *C. arabica* e é o primeiro estado produtor de café. No caso de *C. canephora*, espécie mais adaptada as regiões litorâneas, o estado do Espírito Santo é o estado que mais produz essa espécie. Segundo a Pesquisa Agrícola Municipal (PAM), divulgada pelo Instituto Brasileiro de Geografia e Estatística - IBGE (2010) as principais regiões produtoras de café em 2009 estão apresentadas na Tabela 1.

Tabela 1 Pesquisa Agrícola Municipal (PAM), divulgada pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2010)

<b>Principais Regiões Produtoras de Café no ano de 2009</b>	
Minas Gerais	49%
Espírito Santo	25,4%
São Paulo	8,1%
Bahia	7,2%
Paraná	3,7%
Rondônia	3,8%
Acre, Pará, Ceará, Pernambuco, Rio de Janeiro, Mato Grosso, Mato Grosso do Sul, Goiás e Distrito Federal	2,8%

### **3 ESTRESSE HÍDRICO (ASPECTOS GERAIS)**

O termo estresse é, na maioria das definições, considerado como um desvio significativo das condições ótimas para a vida, e induz a mudanças em todos os níveis funcionais do organismo, as quais são reversíveis a princípio, mas podem se tornar permanentes (LARCHER, 2003).

Dessa forma, o estado hídrico das plantas sofre alterações ao longo do dia, influenciado por processos como absorção de água e transpiração, os quais atingem à maioria dos processos fisiológicos como a fotossíntese, fixação de nitrogênio e respiração. Também afeta diferentes estádios de desenvolvimento da planta, incluindo a germinação de sementes, desenvolvimento de plântulas, crescimento vegetativo e reprodutivo, maturação de sementes e senescência, que são diferencialmente alterados em respostas às condições de estresse (TAIZ; ZEIGER, 2004).

Para evitar danos decorrentes da perda excessiva de água, as plantas promovem o fechamento parcial ou total de seus estômatos. Contudo, quando as plantas são submetidas a uma condição de baixa disponibilidade de água como, por exemplo, durante um período de estresse hídrico o processo de desidratação poderá atingir um estado irreversível, denominado ponto de murcha permanente, podendo levar as plantas à morte por dessecação (-1,5 MPa é o ponto de murcha permanente para o grupo das mesófitas) (TAIZ; ZEIGER, 2004).

Além disso, o estresse hídrico afeta diretamente a quantidade de água no solo, aumentando a resistência ao fluxo de água e conseqüentemente dificultando a absorção da água pelas raízes. Ocorre frequentemente em sistemas agrícolas, sendo um fator limitante na produção de plantas cultivadas. Estima-se uma redução significativa na produtividade, em condições de estresse hídrico (TADA; TAKAOKI, 2009).

Dessa forma as plantas precisam se adaptar às mudanças ambientais desenvolvendo uma série de respostas fisiológicas, bioquímicas e moleculares para sobreviver a essas condições. A base dos processos fisiológicos para essas respostas moleculares é a integração de muitos eventos em uma ampla rede de rotas de sinalização (SHAO et al., 2007).

### **3.1 Respostas ao estresse abiótico em plantas**

A capacidade da planta em manter suas atividades metabólicas enquanto se encontra desidratada é mediada por mecanismos de respostas moleculares e bioquímicas. Mecanismos que são controlados em função do tempo de duração e da intensidade do estresse processam-se nos mais variados níveis de complexidade: morfológicamente, 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 (BRAY, 1993; TAIZ; ZEIGER, 2004).

De acordo com Wang, Vinocur e Altman (2003), para promover a resposta de tolerância ao estresse hídrico é necessário primeiramente a percepção do estresse. Em seguida, há um “disparo” ou início da cascata de transdução de sinais ativando genes específicos, que posteriormente oferecerão respostas bioquímicas à planta. A cascata de eventos moleculares é finalizada em vários níveis de respostas fisiológicas, metabólicas e de desenvolvimento (Figura 3). O mecanismo da transcrição gênica é controlado pela interação de proteínas específicas denominadas fatores de transcrição, que reconhecem seqüências particulares (TAIZ; ZEIGER, 2004). Diferentes genes induzidos pelo

déficit hídrico, alta concentração salina e estresse pelo frio são regulados por rotas de sinalização que ocasionam a ativação desses fatores de transcrição.

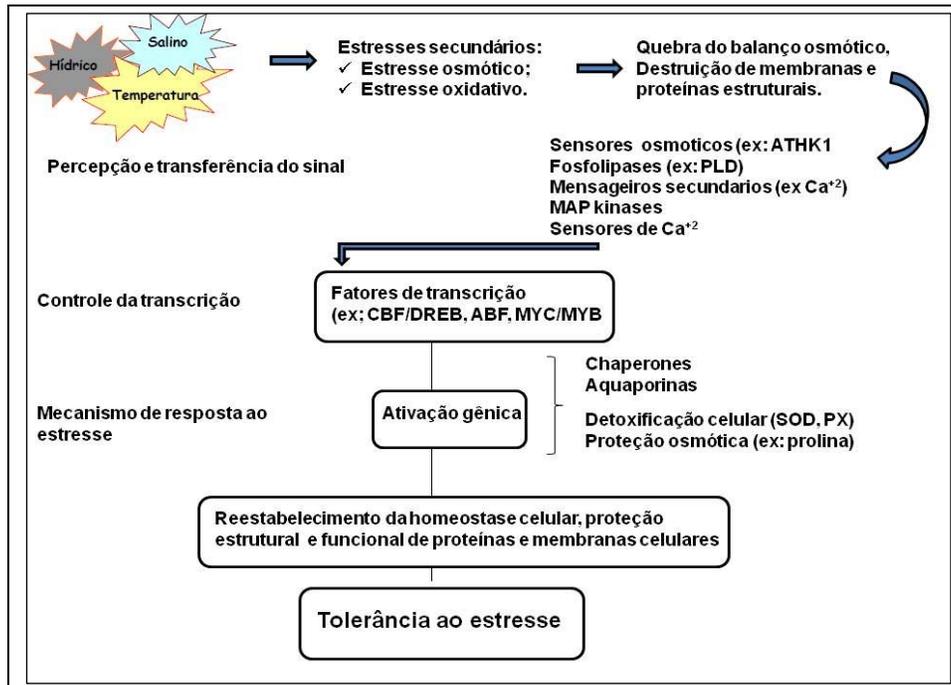


Figura 3 Cadeia de eventos moleculares na resposta das plantas aos estresses abióticos: Percepção e transferência de sinal, controle da transcrição e mecanismos de respostas aos estresses abióticos, adaptado de Wang, Vinocur e Altman (2003)

Condições de estresse hídrico desencadeiam a produção do ácido abscísico (ABA), que causa fechamento dos estômatos e induz a expressão de genes relacionados ao estresse. Muitos genes são induzidos por tratamento com ABA exógeno, enquanto outros não. Isto sugere a existência de vias de sinalização de estresse ABA dependente e ABA independente (Figura 4). Segundo Shinozaki e Yamaguchi-Shinozaki (2007), a hipótese é que existam

seis vias de transdução de sinal para a ativação dos genes induzidos por seca, alta salinidade e frio, sendo três dependentes de ABA e três independentes.

Nesse processo, os genes *DREB* (“Dehydration-responsive element binding protein”) possuem um papel importante. Podem ser divididos em dois grupos de acordo como o estresse envolvido: *DREB1*, relacionado com a tolerância a baixas temperaturas/estresse do frio e *DREB2* associado com a tolerância à falta de água. Todas as proteínas DREB são fatores de transcrição que possuem um domínio altamente conservado chamado ERF/APE2, que é essencial na resposta à expressão do hormônio etileno e a morfogênese floral, respectivamente (LIU, 2007; SHINOZAKI; YAMAGUCHI-SHINOZAKI, 1997).

Nas vias independentes de ABA, as proteínas DREB (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 2005) regulam a expressão de genes induzidos por estresse hídrico, salinidade e frio. Assim, a superexpressão de *DREB1A* (DUBOUZET et al., 2003) em milho transgênico resultou em aumento da tolerância à seca, alta salinidade e congelamento, mostrando que o gene *DREB* pode ser utilizado para aumentar a tolerância de muitas espécies à seca e frio por meio através da engenharia genética (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 2005).

O ABA desempenha uma função vital de sinalização nas plantas em resposta ao estresse, sendo evidenciado pelo estudo de vários genes induzidos pelo estresse hídrico que são também induzidos pelo ABA. Assim, os fatores de transcrição *DREB1/CBFs* (“*DRE-binding protein/C-repeat-binding factor*”) estão envolvidos no controle da expressão gênica em resposta ao frio e as proteínas *DREB2* na resposta a desidratação e alta salinidade. Portanto a terceira via independente de ABA é controlada por estresse hídrico e salinidade, mas não por frio (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2007).

Nesse contexto, existe um grande número de fatores de transcrição (TFs) no genoma vegetal. Muitos destes TFs pertencem a algumas grandes famílias de multígenes *MYB*, *AP2/EREBP*, *bZIP* e *WRKY*. *MYB2* e *MYC2* e *NAC* que também atuam na expressão gênica induzida por ABA às vezes em conjunto com a via de resposta ao ácido jasmônico (JA) e podem estar relacionados à resposta a estresses abióticos. Membros individuais da mesma família podem responder diferentemente a vários estímulos de estresse. Por outro lado, alguns genes de resposta ao estresse podem partir do mesmo TF, indicando a significava sobreposição dos perfis de expressão gênica induzidos em resposta a diferentes estresses (CHEN; MURATA, 2002).

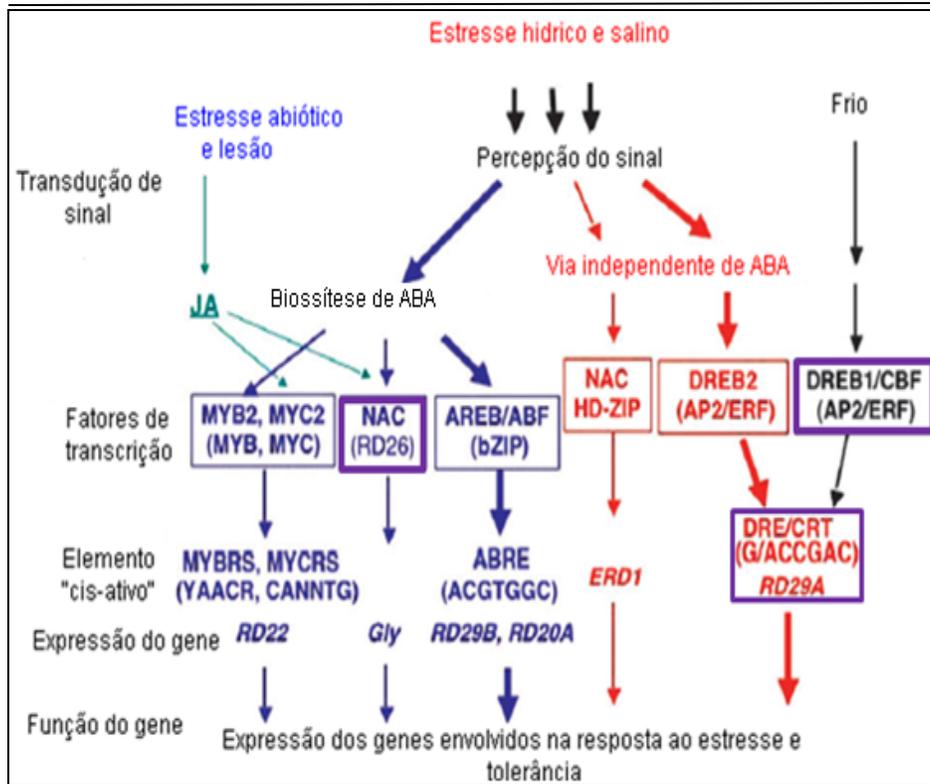


Figura 4 Redes transcripcionais reguladoras de sinais de estresse abiótico e da expressão gênica, adaptado de Shinozaki e Yamaguchi-Shinozaki (2007). Pelo menos seis vias de transdução de sinal existem e estão relacionadas ao estresse hídrico, alta salinidade e ao frio, fornecendo respostas ao estresse: três são dependentes de ABA (importantes na adaptação a desidratação) e três são independentes de ABA. MYB2 e MYC2 são fatores de transcrição (FT) que agem na via de ABA-dependente e possuem a função de ativar a expressão do gene *RD22*. O fator de transcrição NAC RD26 está envolvido na via dependente de ABA. A sequência DRE é importante para a regulação da expressão do gene *RD29A* induzido na resposta ABA independente ao frio e a desidratação

### 3.2 Respostas ao estresse hídrico em *Coffea*

Em consequência da falta de água no solo, o cafeeiro tem seu metabolismo alterado. Com o aumento da severidade do déficit hídrico a fotossíntese é afetada, principalmente devido ao fechamento dos estômatos e ao bloqueio do dióxido de carbono (CO<sub>2</sub>), ocorre também a redução do conteúdo de clorofila em função da desintegração de membranas causada pelo estresse oxidativo em espécies menos tolerantes ao estresse (DAMATTA et al., 2003). Ocorre redução do fluxo de vapor da transpiração, assim como a absorção de água e de nutrientes pelo sistema radicular, especialmente pelas raízes absorventes e, conseqüentemente, diminuição da produção (KUMAR, 1999; MATIELLO; DANTAS, 1987).

Além disso, com a perda do turgor, ocorre uma diminuição do volume celular, um aumento progressivo na concentração do citossol e a desidratação do protoplasto, devido ao fato da célula não ser capaz de retirar água do ambiente (PINHEIRO, 2005).

Contudo, os mecanismos fisiológicos associados à tolerância no gênero *Coffea* parecem estar fortemente relacionados à grande sensibilidade estomática induzida pelo déficit hídrico no solo ou na atmosfera (PINHEIRO, 2005).

Dessa forma, a condutância estomática em *C. arabica* parece constituir um indicador da falta de água, mostrando reduções imediatas tão logo um terço do aporte de água no solo decaia. Ao contrário, *C. canephora* parece exibir um baixo controle estomático da transpiração (DAMATTA et al., 2003). Estudos desenvolvidos por Fahl et al. (2001) demonstraram que o controle da abertura estomática em *C. canephora*, via sinais químicos da raiz, foi menos responsivo à redução da água no solo do que em *C. arabica*. Possivelmente essas variações remontam às origens das duas espécies, *C. arabica* nativo da Etiópia (1600-2800 m de altitude) derivou de uma região com uma estação seca duradoura (4 a 5

meses/ano) e umidade atmosférica relativamente baixa e *C. canephora* é predominante da bacia do Rio Congo (baixa altitude), em típico clima equatorial, com chuvas abundantes e bem distribuídas ao longo de 9-10 meses e umidade próxima da saturação.

Assim, para extrair água do ambiente, as células acionam um mecanismo denominado de ajuste osmótico, que é o processo pelo qual o potencial hídrico pode ser diminuído sem que haja decréscimo no turgor ou no volume celular por meio da acumulação de solutos compatíveis, comumente sintetizados pela célula e que não interferem nas funções das enzimas (TAIZ; ZEIGER, 2004).

Estudos recentes têm investigado a participação do sistema antioxidante em cafeeiros e sua possível correlação com tolerância à seca. Lima (2002) propôs que a tolerância à seca em *C.canephora* poderia estar parcialmente associada ao aumento da atividade de enzimas antioxidantes. Evidências reforçam a existência de uma grande variabilidade genética relativa à tolerância à seca em *Coffea*, inclusive intraespecífica. Estudos com base na caracterização de isoenzimas identificaram 2 grandes grupos em *C. canephora*: o Guineano, originário da Guiné e da Costa do Marfim e o Congolês, com origem na África Central. Neste caso, o grupo de Congolês é dividido em dois subgrupos distintos geneticamente que apresentam também características diferentes em relação à tolerância à seca com o subgrupo (SG1) tolerante e o subgrupo 2 (SG2) sensível a seca (MONTAGNON; LEROY; YAPO, 1992).

### **3.3 Café e biotecnologia**

Por se tratar de uma cultura perene, o melhoramento genético do cafeeiro é lento. A obtenção de um novo cultivar de café a partir do melhoramento convencional demora aproximadamente 30 anos. As plantas

selecionadas são avaliadas, estudadas as diferentes características de interesse são anotadas (BOUHARMONT; AWEMO, 1979; FAZUOLI et al., 2000).

Para garantir o suprimento da variabilidade gênica para os programas de melhoramento do cafeeiro no Brasil, é importante a organização e a conservação de recursos genéticos de café em BAGs. Nesse contexto, BAGs são locais onde são mantidas coleções de indivíduos visando-se preservar a variabilidade genética existente em uma ou mais espécies. Por exemplo, existem os BAGs de *Coffea*, liderados pela Embrapa Café, e executados em parceria com o IAC, IAPAR, INCAPER e EPAMIG (Empresa de Pesquisa Agropecuária de Minas Gerais) e UFV (Universidade Federal de Viçosa). Por meio desses projetos, os recursos genéticos de café são mantidos em diversas instituições de pesquisa de várias regiões cafeeiras do país, diminuindo a possibilidade de perda da variabilidade genética disponível em consequência das adversidades climáticas ou de qualquer outra natureza (INSTITUTO AGRONÔMICO DO PARANÁ - IAPAR, 2009).

Vários genes candidatos (GC) ao estresse biótico e abiótico têm sido identificados em projetos onde genomas inteiros foram sequenciados, como nos casos de *Arabidopsis thaliana* (KAUL et al., 2000) e de arroz (YU et al., 2002). Além disto, os GC podem ser identificados em projetos nos quais foram sequenciados milhares de seqüências oriundas de bibliotecas de cDNA construídas a partir de diferentes órgãos/tecidos e em diferentes condições de estresse, também conhecidos como projetos ESTs (“expressed sequence tags”). Como exemplo o projeto Café da Rede de Sequenciamento financiada pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café, Embrapa e FAPESP (VIEIRA et al., 2006). O Projeto Genoma Café optou pelo sequenciamento de ESTs, onde só os genes expressos foram sequenciados. Utilizando-se essa metodologia, foram obtidas mais de 200 mil seqüências de segmentos de DNA preparados entre outros a partir de folhas, raízes, sementes,

frutos (ANDRADE, 2007; VIEIRA et al., 2006). Entretanto, na fase do genoma funcional, estão agora sendo identificados os genes que determinam e influenciam a qualidade do café, que fornecem um aroma melhor, os genes envolvidos com tolerância as pragas e doenças e também aqueles que podem condicionar ou gerar plantas com maior eficiência no uso dos recursos hídricos (ANDRADE, 2007).

O banco de ESTs oriundo do projeto Genoma Café esta sendo utilizado por diferentes pesquisadores para: a) desenvolver novos e eficientes marcadores moleculares, que podem ser utilizados para análise de diversidade genética nos programas de melhoramento, b) identificar e localizar genes de interesse agrônômico, c) estudar a função de genes dentro de vias metabólicas, d) estudar a regulação de genes em resposta a diferentes estresses, e) clonar genes, permitindo incluir a engenharia genética entre as estratégias do melhoramento, g) ampliar a base de conhecimento acerca dos processos metabólicos do cafeeiro (CAIXETA et al., 2008; LASHERMES; ANDRADE; ETIENNE, 2008).

### **3.4 Marcadores moleculares**

Cada cromossomo contém uma longa e única molécula de DNA, além de proteínas que atuam no empacotamento desta molécula. As tecnologias de análise molecular da variabilidade do DNA permitem determinar pontos de referência nos cromossomos denominados “marcadores moleculares”. Esses marcadores oferecem aos melhoristas a possibilidade de acessar o genótipo da planta ao invés de simplesmente o fenótipo. Em razão do fenótipo ser a expressão do genótipo, sob condições ambientais específicas, este pode mudar com o ambiente (KUMAR, 1999).

Marcadores de DNA têm sido utilizados para análises de divergência genética (CATTANEO, 2001), mapeamento genético (RAFALSKI, 2002),

filogenia (GEHRIG; ROSICKER; KLUGE, 1997), melhoramento genético visando resistência a estresses bióticos e abióticos (SANTOS, 2000). Todos estes estudos buscam identificar variações em sequências genômicas que possam estar relacionadas com um fenótipo observado. Podem, também, minimizar os problemas associados aos métodos de melhoramento clássico, como o enorme trabalho e tempo, principalmente, em se tratando de culturas perenes.

Os primeiros trabalhos que utilizaram marcadores moleculares para o estudo da diversidade genética foram realizados em *C. arabica* com o uso de marcadores RAPD (LASHERMES, 1996), e um dos primeiros estudos em *C. canephora* foi realizado por Dusset et al. (1999), com marcadores que permitem detectar as diferenças entre indivíduos diretamente no DNA, denominados RFLP (“Restriction Fragment Length Polymorphism”) para polimorfismo no fragmento de restrição (BORÉM; CAIXETA, 2009).

Os marcadores moleculares diferenciam-se pela tecnologia utilizada para revelar a variabilidade em nível do DNA, distinguem-se quanto à capacidade de gerar diferenças entre indivíduos (polimorfismo), custo, facilidade de uso, e reprodutibilidade (BORÉM; CAIXETA, 2009).

Entre as técnicas de detecção de polimorfismos estão os marcadores de polimorfismo de nucleotídeos únicos (SNPs) e os microssatélites (SSR: “Simple sequence repeat”), marcadores altamente polimórficos, bastante empregados em estudos de polimorfismos entre sequências e em estudos de tolerância a estresses bióticos e abióticos. Essa técnica permite a obtenção de muitos marcadores moleculares cobrindo todo o genoma do organismo (SANTOS et al., 2005).

### **3.5 SNPs (Single Nucleotide Polymorphisms, INDELs – Insertion/Deletions)**

Os polimorfismos de modificações nucleotídicas se tornaram a principal escolha de marcadores devido à alta frequência no genoma. Esses polimorfismos

podem ser detectados com a análise de ESTs disponíveis (análises *in silico*), nas seqüências de GC (análise *in vivo*) ou nas seqüências de DNA genômico (gDNA). Os SNPs mais comuns são os de transição, onde uma purina é substituída por outra purina (A <-> G) ou uma pirimidina por outra pirimidina (C <-> T), e de transversão em que uma base purínica é substituída por uma pirimídica ou vice-versa. Também são considerados polimorfismos de nucleotídeos a inserção ou ausência de nucleotídeos em uma seqüência, chamados de *INDELS* (“*Insertion / Deletions*”).

Os SNPs são as formas mais abundantes de variação do genoma. Eles podem ser utilizados na análise de diversidade genética e caracterização de recursos genéticos, análises filogenéticas, diagnóstico e identificação de cultivares e construção de mapas com alta densidade de marcadores e no melhoramento a partir de seleção assistida por marcadores (RAFALSKI, 2002). A seleção assistida por marcadores (MAS) fornece a oportunidade de se acelerar os processos de desenvolvimento de novos cultivares de café (ANDRADE, 2007). Entretanto, marcadores convencionais, tais como os polimorfismos de fragmentos de restrição (RFLPs), os DNA polimórficos amplificados ao acaso (RAPDs), os polimorfismos amplificados a partir do tamanho do fragmento (AFLPs) e seqüências de repetições simples (SSRs), geralmente não são desenvolvidos na seqüência codante dos genes de interesse. Ao contrário, marcadores funcionais tais como os SNPs são geralmente desenhados com base nos polimorfismos dentro das regiões transcritas de genes e conseqüentemente correlacionados com a função desses genes (ANDERSEN; LÜBBERSTEDT, 2003). Assim, os SNPs que alteram o aminoácido codificado, são chamados não-sinônimos, e podem trazer modificações estruturais e funcionais nas proteínas. Entretanto, SNPs nas regiões codantes que não alteram o aminoácido codificado e são classificados como sinônimos, enquanto os SNPs das regiões não codantes são chamados de anônimos (RAFALSKI, 2002).

Recentemente um trabalho sobre análise da diversidade nucleotídica de genes envolvidos na biossíntese de sacarose e diterpenos em café, demonstrou a importância da utilização de ESTs para identificação de polimorfismos (YANAGUI et al., 2009). Os resultados foram comparados com a análise *in silico* dos EST disponíveis, para verificar se os genes são submetidos às mesmas pressões de seleção e se apresentam os mesmos níveis de diversidade.

O número e distribuição de SNPs são influenciados pela pressão de seleção. Além disso, quando essas mutações ocorrem dentro de um gene, podem alterar a formação da respectiva proteína. Mas os SNPs também podem ser encontrados fora das sequências codantes, tal como, no promotor (BARKER et al., 2003). Se bem conhecidos e mapeados, os SNPs podem ser utilizados para avaliar a diversidade nucleotídica de determinada cultura, desde que se tenha o prévio conhecimento das sequências (RAFALSKI, 2002). Após a identificação dos SNPs é necessário validá-los, ou seja, determinar a frequência do SNP em um grupo de indivíduos, certificando-se de sua natureza polimórfica (BARKER et al., 2003).

A quantidade de informação gerada pelos projetos genomas nos últimos anos, como os ESTs e as ferramentas de bioinformática disponíveis, propicia a identificação de SNPs e de novos marcadores microssatélites (SSR). O uso de ESTs para a identificação de SNPs tem sido utilizado em várias espécies de importância agrícola, como por exemplo, milho (USECHE, 2001), cana-de-açúcar (GRIVET et al., 2003), trigo (SOMERS, 2003) e café (YANAGUI et al., 2009).

## **4 MATERIAL E MÉTODO**

### **4.1 Validação da expressão de genes candidatos em cultivares de *C. arabica* cultivadas no campo**

Mudas dos cultivares Rubi de *C. arabica* (cultivar considerado sensível à seca) e Iapar59 (I59: cultivar considerado tolerante a seca) foram plantadas em dezembro de 2007 no campo experimental da Embrapa Cerrados – DF (CPAC) e utilizadas para a validação da expressão dos genes candidatos. Exemplares de ambos cultivares foram submetidos à irrigação e não irrigação durante todo o período de 2008 e 2009. As plantas foram divididas em blocos de 1 até 6 (P1 à P6), com três pontos de colheita ao longo de cada ano (Figura 5). Em cada ponto de colheita foram colhidas raízes, folhas e ramos de plantas (três repetições biológicas) de todos os cultivares nas condições (I: Irrigadas, NI: não irrigadas e NI/I teste de “recuperação” com plantas não irrigadas em 2008 e irrigadas em 2009). Neste trabalho utilizamos preferencialmente as folhas dos cultivares colhidas nos pontos de colheita P1, P2, P4 e P5 para a análise de expressão por meio da técnica de PCR quantitativo em tempo real (qPCR) (7500 software v2.0.1).

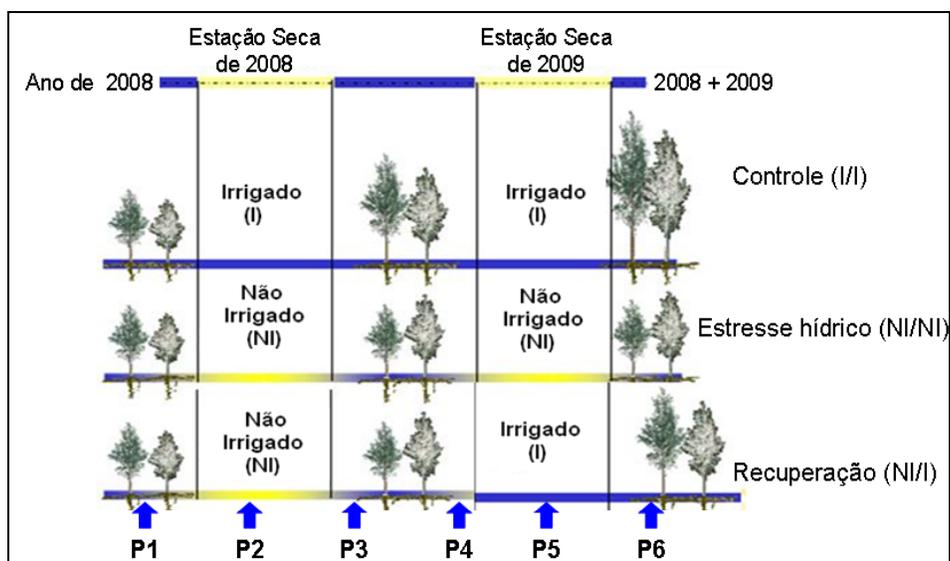


Figura 5 O ponto de colheita (P1) corresponde a plantas jovens, colhidas no final do período de chuva (maio 2008). O ponto de colheita (P2) corresponde às plantas colhidas no período de seca (agosto 2008). O ponto de colheita (P3) corresponde à plantas controles (período de chuva após o período de estresse Novembro 2008). O mesmo esquema foi repetido em 2009 (P4: abril 2009, P5 agosto 2009 e P6 fevereiro de 2010)

#### 4.1.1 Caracterização fisiológica

As taxas de transpiração, de fotossíntese e de crescimento são afetadas pelas alterações no estado hídrico das plantas. Um dos meios para se caracterizar o estado hídrico das plantas é a avaliação do potencial de água (ANGELOCCI, 2002).

Nesse contexto, as análises de fisiologia foram baseadas nas medidas do potencial hídrico de antemanhã ( $\Psi_{am}$ ) das plantas dos cultivares Rubi e I59 de *C. arabica*, irrigado (I) e não irrigado (NI). O potencial de água nas folhas dos cafeeiros foi determinado entre 4:00 e 5:00 horas, uma vez por semana, utilizando-se uma bomba de pressão tipo Scholander. Para se medir o potencial

hídrico, foram coletadas três folhas de cada planta, totalizando nove leituras por cultivar. Foram utilizadas para as medidas somente folhas totalmente expandidas e não danificadas, localizadas no terceiro par a partir da extremidade dos ramos plagiotrópicos.

#### **4.1.2 Extração de RNA**

Folhas de cafeeiros dos cultivares de *C. arabica* cv. Rubi e I59 foram coletadas, entre as 10:00 e 12:00 horas, acondicionadas em tubos falcon, congeladas em N<sub>2</sub> líquido e armazenadas a -80 °C, para posterior extração de RNA total. Todos os materiais utilizados para extração (cadinhos, pistilos de porcelana, microtubos, ponteiros e água destilada) foram autoclavados para limitar a ausência de atividade RNase e garantir a qualidade do RNAs extraídos.

As paredes celulares foram rompidas com o objetivo de liberar os constituintes celulares e as cadeias de ácidos nucleicos. Essa etapa foi realizada com o congelamento do tecido vegetal em N<sub>2</sub> líquido e posterior quebra mecânica, pulverizando as folhas com o auxílio de um pistilo e um almofariz. Após a pulverização do tecido vegetal com N<sub>2</sub> líquido, aproximadamente 10 mg do material pulverizado foi transferido para um microtubo Eppendorf (1,5 ml). A extração dos RNAs foi realizada utilizando-se o tampão de extração CONCERT® (Invitrogen) com a adição de 500 µL de tampão no tecido pulverizado e posterior agitação com vórtex para a homogeneização. Os microtubos foram mantidos na horizontal por 5min. à temperatura ambiente. Então, foram centrifugados por 2min. aos 4 °C em 16100 x g. A fase sólida contendo componentes de parede celular foi descartada e a fase líquida foi transferida para novos microtubos contendo 100 µL de 5M NaCl e 300 µL de clorofórmio para permitir a separação das moléculas proteicas e açúcares das cadeias nucleotídicas. Após nova homogeneização com vórtex, os microtubos

foram centrifugados por 10 min. aos 4 °C por 16100 x g, o que possibilitou coletar a fase superior que continha os ácidos nucleicos em solução. A precipitação dos RNAs totais foi realizada com a adição de 500 µL de Isopropanol, a temperatura ambiente durante 30min. Seguida de uma nova centrifugação de 15min. aos 4 °C por 16100 x g. O pélete foi lavado com 600 µL de etanol 70% para retirar o excesso de sais, colocado em banho seco a 37 °C até secar e ressuscitado em 20 µL de água Milli-Q autoclavada. Cada extração de RNA foi realizada em triplicata.

#### **4.1.3 Análise da qualidade e quantificação das alíquotas de RNA**

Para se avaliar a integridade das amostras extraídas, o RNA foi submetido à eletroforese em gel de agarose 0,8% corado com 0,7 µL de Brometo de Etídio ( $0,5 \mu\text{g.mL}^{-1}$ ), posteriormente visualizado sob luz ultravioleta, bandas individuais de RNA ribossomais 26S e 18S e a imagem foi captada pelo fotodocumentador da Loccus Biotecnologia. As amostras foram quantificadas em espectrofotômetro (Nanodrop® Espectrophotometer ND-1000). A qualidade foi avaliada pelo espectro (220-600nm) com a razão  $DO_{260}/DO_{280}$ . Se a razão for menor que 1,8, pode indicar a presença de contaminação de proteínas nas amostras de RNA.

#### **4.1.4 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 no termociclador modelo PTC-100 (MJ Research). Ao microtubo contendo 8 µg de RNA total foram adicionados 1 µL de tampão 10X

(presente no kit) e 1  $\mu\text{L}$  da enzima RQ1 RNase Free. Essa reação foi incubada aos 37 °C por 30 min. para ativação da enzima. Para inativar a enzima foi adicionado 1  $\mu\text{L}$  de RQ1 RNase Stop Solution (EGTA 20 mM pH 8) e a reação foi incubada aos 65 °C por 10min. 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.

#### **4.1.5 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 dT<sub>15</sub>, o kit SuperScript™ II Reverse Transcriptase (Invitrogen) e o termociclador modelo PTC-100 (MJ Research). Em microtubos estéreis, adicionou-se 5  $\mu\text{g}$  de RNA total, 1  $\mu\text{L}$  de oligo- dT<sub>15</sub> concentração final de 10  $\mu\text{M}$  e 1  $\mu\text{L}$  do mix de dNTP na concentração final de 0.5 mM cada dNTP. A mistura foi aquecida aos 65 °C por 5min. para desnaturação das estruturas secundárias e para o anelamento dos primers ao RNA. Após serem mantidas em gelo por 5min. acrescentou-se ao mesmo tubo 4  $\mu\text{L}$  do tampão First-Strand 5X, 1  $\mu\text{M}$  de DTT (0,1M) e 1  $\mu\text{L}$  de RNase Out (40U.  $\mu\text{L}^{-1}$ ). Os tubos foram incubados no termociclador, aos 42 °C, durante 2 min. Foi adicionado 1  $\mu\text{L}$  da enzima SuperScript II™ 200U. $\mu\text{L}^{-1}$  e água Milli-Q para se completar o volume final de 20  $\mu\text{L}$ . A reação foi incubada aos 42 °C por 50min. para a síntese da primeira fita de cDNA, sendo posteriormente incubada aos 72 °C durante 15min. para inativar a enzima. As amostras de cDNA foram armazenadas aos -20 °C.

#### 4.1.6 Desenho de primers para as análises de expressão pela qPCR

Os *primers* utilizados para se analisar a expressão dos GC (Tabela 2) em cultivares de *C. arabica* cultivados no campo, foram desenhados utilizando-se o programa “Primer Express 3.0” (Applied Biosystems). 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. *Primers* específicos para os GC foram desenhados a partir da sequência dos *Contigs* do Projeto Genoma Café encontrados na base de dados da Embrapa por meio do endereço eletrônico <https://alanine.cenargen.embrapa.br/cafEST>. Para o gene *RBCS* que codifica a pequena subunidade da RUBISCO foram usados dois pares de *primers* (*RBCS1A* e *RBCS1B*) para se testar a expressão de diferentes isoformas (Gráfico 5C a F).

O gene *GAPDH* que codifica para a enzima gliceraldeído 3-fosfato desidrogenase (EC 1.2.1.12) foi usado como controle endógeno constitutivo da expressão. Os *primers* foram sintetizados pela empresa *Eurofins MWG Synthesis GmbH* e ressuspensos em água ultra pura para manutenção em solução estoque a 100 $\mu$ M.

Tabela 2 *Primers* utilizados para análise de qPCR. As sequências dos *primers* e o tamanho dos amplicons amplificados estão indicados na tabela

<i>Primers</i>	Sequências dos <i>Primers</i>	Amplicon (pb) <sup>a</sup>
<i>DREBA09-F</i>	5'CAATGCCTGCAAAGCCAATTA3'	80
<i>DREBA09-R</i>	5'TTTTCCTGCCTGCACGTTTC3'	
<i>RD29-F</i>	5'TCCCCAGCGGAGTATGCA3'	80
<i>RD29-R</i>	5'GCATCTGCGACTTCTGGTAAA3'	
<i>NACRD26-F</i>	5'TTTGGCCCTGCGCTCTAGT3'	100
<i>NACRD26-R</i>	5'AAGCGGGTCAGTTTCTCGAA3'	
<i>GC10-F</i>	5'TAGCCTTGTTCTTTTAGGGAGTCTTATC3'	134
<i>GC10-R</i>	5'AGAGCTTCGTCCAGGAAGAAGA3'	
<i>MYB61-F</i>	5'CTCCTTGCTGTGAGAAAAGCTCAT3'	100
<i>MYB61-R</i>	5'CAGCAGCCTTCCCCATGT3'	
<i>CCOAMT-F</i>	5'GGCTGGTGTGTCCATAAAAATT3'	100
<i>CCOAMT-R</i>	5'GAAATCAAAAATTCCATGGTTCTTG3'	
<i>AC1-F</i>	5'AAGGCCATTGTCGGACTTCA3'	100
<i>AC1-R</i>	5'TTGTTTGCAACTCTGCAGTGATT3'	
<i>OEC-F</i>	5'CAGGGCAATCAAGGTTGGA3'	102
<i>OEC-R</i>	5'CGGTCCTTGAGTGGCAAATC3'	
<i>M6FR-F</i>	5'ACGAGTAAAGGGCTGGCTAAGA3'	143
<i>M6FR-R</i>	5'TTCCAAACGTCCATCCCTT3'	
<i>RBCS1B-F</i>	5'CCGTCCTCTTCCCCTCAAAT3'	91
<i>RBCS1B-R</i>	5'CCTGAAAAGTACAGCCCCAGTTC3'	
<i>RBCS1A-F</i>	5'CCATGTGGAAGCTGCCTATGTT3'	189
<i>RBCS1A-R</i>	5'GAATTTGTGAAGAAGGGGCTTAAA3'	
<i>GAPDH-F</i>	5'TTGAAGGGCGGTGCAA3'	59
<i>GAPDH-R</i>	5'AACATGGGTGCATCCTTGCT3'	

#### 4.1.7 Análises de expressão pela PCR quantitativo em tempo real (qPCR)

A técnica utilizada para validação dos genes candidatos foi o PCR quantitativo em tempo real (qPCR). Os procedimentos de PCR em tempo real foram conduzidos utilizando-se o equipamento 7500 Fast Real-Time PCR System, seguindo as recomendações do fornecedor (Applied Biosystems). A reação foi realizada com a utilização do fluoróforo SYBR<sup>®</sup> Green (Invitrogen), o

qual se baseia na capacidade do fluoróforo de intercalar as fitas duplas de cDNA. O cDNA foi obtido a partir de RNA extraído de tecidos de folhas de *C. arabica* como descrito em 5.1.3.

#### **4.1.7.1 Avaliação da eficiência dos primers**

Os pares de *primers* usados para as análises de qPCR são descritos na Tabela 2. 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 (*GAPDH*).

Após determinar a melhor diluição do cDNA, foram realizadas quatro diluições seriadas do concentrado de cDNA ( $120\text{ng}\cdot\mu\text{L}^{-1}$ ) (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 88% e 98%, encontrando-se próximos daqueles aceitáveis ( $100 \pm 10\%$ ) (manual da Applied Biosystems).

#### **4.1.7.2 Análises da expressão por PCR quantitativo em tempo real**

As amostras foram processadas em triplicatas, sempre acompanhadas de um controle negativo (NTC: “*no template control*”) que não contém cDNA. O controle negativo nas reações é usado para verificar a ausência de contaminação de cDNA nos mixes de PCR.

Com o objetivo de se verificar a especificidade de anelamento dos *primers* aos fragmentos 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 GAPDH, considerado controle endógeno eficiente para café (BARSALOBRES-CAVALLARI et al., 2009). A normalização foi realizada utilizando-se a equação  $\Delta C_T = C_T$  (gene alvo) -  $C_T$  (controle endógeno). A calibração foi determinada pela fórmula  $\Delta\Delta C_T = \Delta C_T$  (amostra) -  $\Delta C_T$  (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 dados foram analisados no programa 7500 Fast Software (software v2.0.1). Para cada reação, foram utilizadas diluições adequadas de cDNA com 0,4  $\mu$ L de todos os pares de *primers* 10 mM e 5  $\mu$ L de 2xMaster Mix SYBR green (Invitrogen). A reação foi completada com 3,6  $\mu$ L de água Milli-Q para um volume final de 10  $\mu$ L para cada amostra. As amostras foram realizadas em triplicatas. Os resultados foram normalizados usando  $C_T$ s (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).

## 4.2 Busca de SNPs

Amostras de DNAs genômicos (gDNA) de diferentes espécies e cultivares de *Coffea* foram fornecidas pelo IAPAR e pelo CIRAD (Tabela 3). Assim encontram-se, plantas de *C. arabica* da Etiópia (grupos ET1 até 4), cultivares comerciais de *C. arabica* sem introgressão recente de *C. canephora* (Typica, Bourbon, Mundo Novo e, Catuai), cultivares com introgressão de *C. canephora* obtidos por meio de programas de melhoramento que usaram o

Híbrido de Timor, tal como o cultivar Iapar 59 (I59), acessos de *C. canephora* representando a diversidade genética dessa espécie como plantas dos grupos B, C e sub-grupos SG1 e SG2 de congolês e os clones 14 e 22 de Conilon foram usados para identificar os GC, e também espécies geneticamente mais distantes como *C. racemosa* e *C. eugenioides* (LASHERMES et al., 1999).

Tabela 3 Lista de gDNA dos diferentes genótipos. Espécies e sub grupos de café usados para busca e análise dos SNPs para os genes candidatos *DREB*, *RD29* e *NAC-RD26*

Genótipos	Espécie	Subgrupo
01- Mundo novo	<i>C. arabica</i>	Comercial
02- Catuai 25	<i>C. arabica</i>	Comercial
03- Typica	<i>C. arabica</i>	Comercial
04- Bourbon	<i>C. arabica</i>	Comercial
05- Rubi	<i>C. arabica</i>	Comercial
06- IAPAR 59	<i>C. arabica</i>	Comercial
07- E516	<i>C. arabica</i>	Etiópia 1
08- E464	<i>C. arabica</i>	Etiópia 1
09- E007	<i>C. arabica</i>	Etiópia 2
10- E237	<i>C. arabica</i>	Etiópia 2
11- E017	<i>C. arabica</i>	Etiópia 3
12- E238	<i>C. arabica</i>	Etiópia 4
13- E123B	<i>C. arabica</i>	Etiópia 4
14- E123A	<i>C. arabica</i>	Etiópia 4
15- UW099	<i>C. canephora</i>	Uganda Selvagem (região Itwara)
16- UW002	<i>C. canephora</i>	Uganda Selvagem (região Kibale)
18- C3001	<i>C. canephora</i>	Congolês SG1
19- Canephora (G21)	<i>C. canephora</i>	Congolês SG1
20- G2011	<i>C. canephora</i>	Congolês SG2
21- C1007	<i>C. canephora</i>	Congolês B
22- C4001	<i>C. canephora</i>	Congolês C
23- Clone 14	<i>C. canephora</i>	Comercial
24- Clone 22	<i>C. canephora</i>	Comercial
25- G2020	<i>C. canephora</i>	Guine
26- Psilanthus	<i>bengalensis</i>	P. bengalensis
27- Racemosa	<i>C. racemosa</i>	Racemosa
28- Eugenioides	<i>C. eugenioides</i>	Eugenioides

#### 4.2.1 Desenho de *primers* para amplificação dos genes utilizados na busca de SNPs

Vários pares de *primers* foram desenhados por meio do programa “Primer Express” (Applied Biosystems) nas extremidades 5’ e 3’ dos Contigs identificados na base de dados da Embrapa (<https://alanine.cenargen.embrapa.br/cafEST>), acesso restrito a usuários cadastrados. Para amplificar esses genes, várias combinações de *primers* foram testadas (resultados não apresentados) e somente os pares *primers* que apresentaram um produto de PCR específico e de tamanho esperado em relação ao tamanho dos Contigs correspondentes foram usados (Tabela 4). O desenho dos *primers* foi feito para se amplificar segmentos de gDNA de tamanho igual ou maior ao tamanho das sequências de cDNA dos genes.

Para se avaliar a existência de diferentes formas alélicas para cada GC, foram adicionados nas extremidades 5’ dos pares de *primers* adaptadores M13For ou M13Rev que permitiram o sequenciamento direto dos produtos de PCR amplificados sem clonagem. Para se identificar as formas alélicas presentes em cada genoma, *primers* (sem os adaptadores M13) foram usados para amplificar e sequenciar as mesmas porções de gDNA, posteriormente as clonagens em *E. coli*.

Tabela 4 *Primers* utilizados para as ampliações de gDNA para as buscas dos SNPs. Onde <sup>(1)</sup> e <sup>(2)</sup> corresponde a sequência descrita mais os adaptadores M13 (-21) e M13 (-40) respectivamente, utilizados para a realização do sequenciamento dos produtos de PCR dos GCs clonados vetor plasmidial e transfecção

Primers	Sequências dos Primers	Amplicon (pb)
<i>DREB1A-F</i> <sup>(1)</sup>	5'GTTGAATTAACCTCCTCACTGTCCACTA3'	877
<i>DREB1A-R</i> <sup>(2)</sup>	5'CCAAAAACTGCAGTACGGAATAGA3'	
<i>RD29-F</i> <sup>(1)</sup>	5'TTGTTGAAACAGGCATGGAATC3'	1707
<i>RD29-R</i> <sup>(2)</sup>	5'GAGCAAAAAGAATGGGAAAATCCT3'	
<i>NACRD26-F</i> <sup>(1)</sup>	5'AAAAAAAATGGGTGTTTCGAGAAACT3'	1082
<i>NACRD26-R</i> <sup>(2)</sup>	5'CTACTCTTCATGGCCTAAAACCCAT3'	
<sup>(1)</sup> M13 (-21)	5'TGTA AACGACGGCCAGT3'	
<sup>(2)</sup> M13 (-40)	5'GTTTCCAGTCACGAC31	

#### 4.2.2 Reação em cadeia da polimerase (PCR) e amplificação do gDNA

Para a amplificação dos fragmentos foi utilizado 6.25 ng de gDNA de cada cultivar acima citado. Foram realizadas duas reações de PCR:

1<sup>a</sup> Com o objetivo de sequenciar produtos purificados após e diretamente da reação de PCR usando os pares de *primers* com adaptadores M13For e M13Rev.

2<sup>a</sup> Com o objetivo de sequenciar para cada GC produtos de PCR independentes e clonados em *E. coli*. Como o *C. arabica* é alotetraploide e o *C. canephora* diploide, os GC podem ter várias formas alélicas. Para se identificar todas essas isoformas, decidimos sequenciar 10 clones para cada GC, com o intuito de se sequenciar o máximo possível dos alelos. As sequências resultantes foram alinhadas, pelo emprego de aplicativos computacionais visando à identificação de regiões polimórficas.

Independentemente da estratégia seguida, as reações de PCR foram realizadas com um volume final de 50 µL utilizando-se a enzima de alta

fidelidade (Platinum *Taq* DNA Polimerase Hight Fidelity, Invitrogen), que possui atividade exonuclease 3'- 5', necessária para correção de erros de polimerização no DNA neo-sintetizado. Assim erros de síntese não são confundidos com os SNP verdadeiros. Todas as amplificações foram realizadas conforme o protocolo do fabricante. As reações foram incubadas no termociclador modelo PTC-100 (MJ Research) utilizando-se os seguintes parâmetros: desnaturação inicial de 94 °C por 2 min., seguidos de 40 ciclos de amplificação compostos por 3 etapas: 94 °C (desnaturação), por 30 seg.; 55 °C (anelamento), por 30 seg.; 72 °C (extensão), por 4 min. e de uma etapa final de polimerização (10 min. aos 72 °C). Após as reações de PCR os fragmentos foram analisados quanto à integridade por meio de eletroforese em gel de agarose, como descrito previamente. Todos os produtos foram sequenciados em ambas as direções, quer por seqüenciamento direto ou após clonagem (Figura 6). As sequências resultantes foram alinhadas, pelo emprego de aplicativos computacionais visando à identificação de regiões polimórficas.

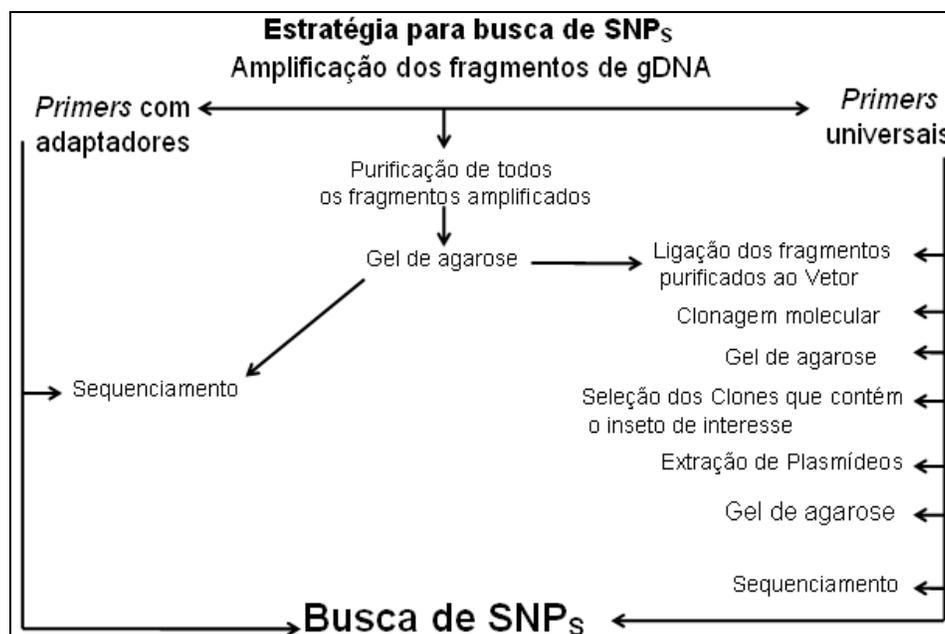


Figura 6 Estratégia para busca de SNPs. Os fragmentos de gDNA foram amplificados e purificados, para retirar restos de *primers*, tampões e amplificações não específicas. Os produtos das amplificações realizadas com os *primers* que possuem adaptadores foram sequenciados logo após as purificações. Os produtos das amplificações realizadas com os *primers* que não possuem adaptadores foram purificados, visualizados em gel de agarose, ligados ao vetor (pCR®4.0Kb-Topo). A clonagem foi realizada e novo gel de agarose foi feito após a reação de PCR sobre as colônias para se identificar os clones que possuem os insertos de interesses. Procedeu-se a extração dos plasmídios que foram posteriormente submetidos ao sequenciamento

#### 4.2.3 Purificação de fragmentos de gDNA a partir do gel de agarose

Todos os produtos de PCR foram purificados, para se eliminar qualquer vestígio de contaminação e resíduos das reações, como por exemplo, restos de *primers* e de dNTPs. Os amplicons a serem recuperados foram aplicados em gel de agarose 0,8% usando-se o tampão de corrida TAE 1X. As bandas foram

visualizadas sob luz ultravioleta “transiluminador”, as porções do gel contendo os fragmentos de interesse foram recortadas e transferidas para microtubos com o auxílio de uma lâmina de bisturi nova. O isolamento dos fragmentos de DNA a partir do gel de agarose foi feito utilizando-se o kit “Wizard SV Gel and PCR Clean-Up System” (Promega), seguindo as normas do fabricante. Após purificados os fragmentos de DNA foram novamente visualizados em gel de agarose para a confirmação que não ocorreu perda ou degradação de material durante a purificação dos mesmos.

#### **4.2.4 Ligação dos fragmentos purificados e clonagem**

Após a purificação com o Kit Wizard, os amplicons foram submetidos à clonagem no vetor pCR<sup>®</sup>II usando-se TOPO TA Cloning<sup>®</sup> Kit Dual Promoter (Invitrogen) (Figura 7). As reações de ligação dos fragmentos ao vetor foram realizadas a temperatura ambiente (22-23 °C) por 20 min., misturando-se 2 µL do produto da PCR purificado, 0,5 µL de solução salina (salt solution 1.2M NaCl e 0.06M MgCl<sub>2</sub>) “especificações do fabricante” e 0,5 µL de vetor. As misturas foram colocadas no gelo para posterior utilização na transfecção bacteriana.

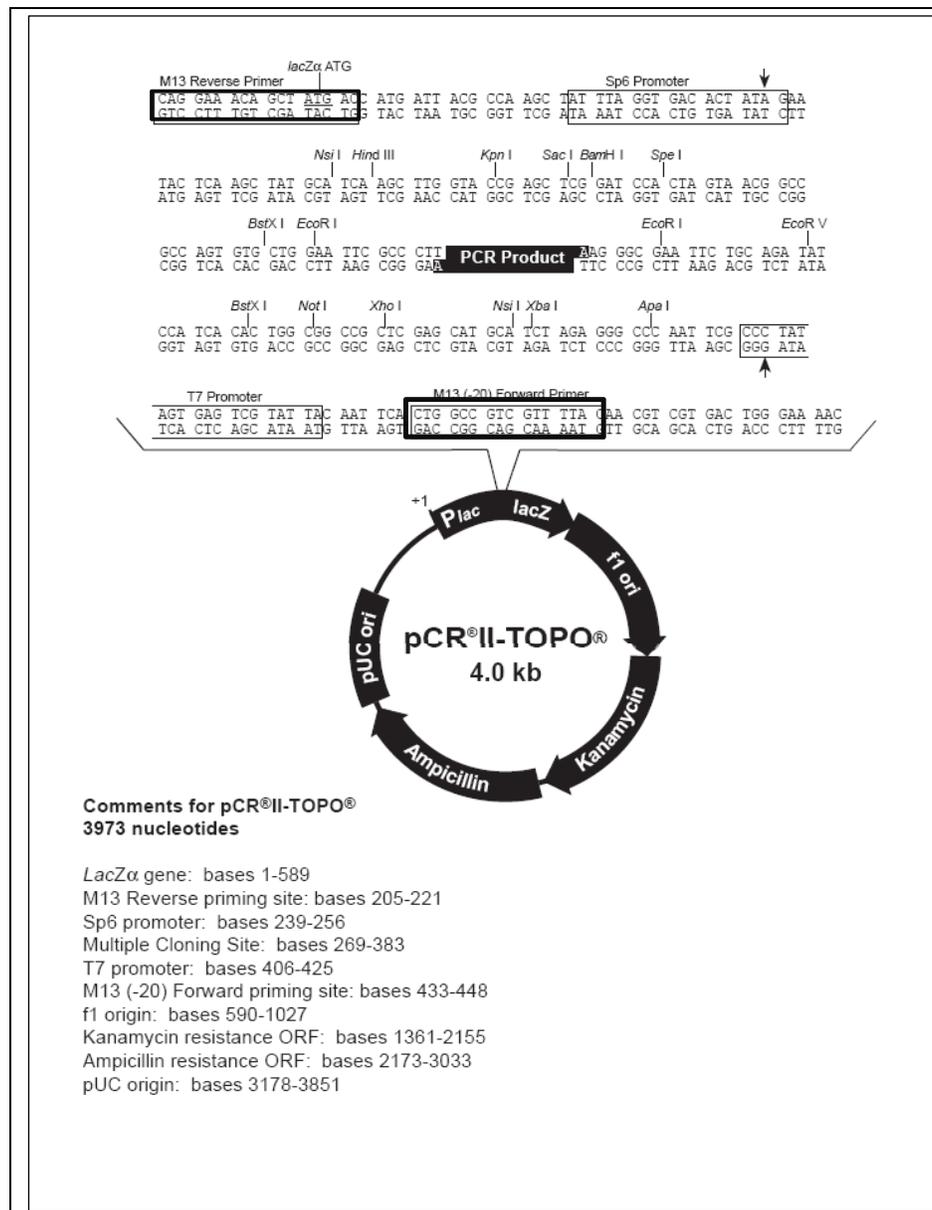


Figura 7 Mapa do vetor pCR®II TOPO TA Cloning (Invitrogen)  
Fonte: Manual Invitrogen

#### 4.2.5 Transfecção bacteriana

Uma alíquota de 100  $\mu\text{L}$  da preparação de célula competente (*E. coli* DH5 $\alpha$ ) foi acondicionada em banho de gelo e misturada com 2 a 5  $\mu\text{L}$  da ligação. A mistura foi incubada em gelo por 30 min. adicionais e levada ao termociclador (modelo PTC-100, MJ Research) para receber o choque térmico (42 °C por 1 min.). Após o choque térmico, a mistura de células/DNA foi incubada aos 4 °C por 2 min., resuspendida em 600 mL de meio SOC pré incubado aos 37 °C, e incubada durante 60 min. aos 37 °C com agitação (300 rpm, modelo do agitador horizontal). A seleção dos clones transformados foi realizada em meio sólido (LB Ágar), contendo ampicilina, X-Gal e IPTG nas concentrações de 100  $\mu\text{g.mL}^{-1}$ , 40  $\mu\text{g.mL}^{-1}$  e 24  $\mu\text{g.mL}^{-1}$ , respectivamente, onde foram espalhados 150  $\mu\text{L}$  da solução de cultura bacteriana. Após 16 horas de incubação em estufa as colônias que apresentarem coloração branca foram selecionadas e estriadas em uma nova placa de meio LB sólido, contendo ampicilina, X-Gal e IPTG nas concentrações indicadas previamente e incubadas novamente em estufa aos 37 °C por 16 horas.

#### 4.2.6 Extração de DNA plasmidial “Miniprep”

As extrações do DNA plasmidial das colônias selecionadas foram realizadas em placas de 96 poços “Deep well” (Axigen). Após verificada a uniformidade do crescimento das colônias, a placa “Deep well” foi centrifugada em 3220 g por 6 min. para peletizar células bacterianas. O adesivo foi retirado e o sobrenadante descartado, a placa ficou invertida sobre papel toalha por alguns minutos para escorrer o excesso de meio. Foram adicionados 240  $\mu\text{L}$  de solução I (GET) em cada poço da placa. A placa foi selada, agitada em vórtex até que as células fossem ressuspendidas. A placa foi centrifugada a 3220g por 6 min

20°C. Após a centrifugação, o sobrenadante foi descartado e foram adicionados 80 µL de solução I em cada poço da placa. A placa foi agitada em vórtex para homogeneização, 60 µL da suspensão de células foram transferidos para uma placa com fundo U (Axigen). Em cada poço, foram adicionados 1 µL de RNase (10 mg.mL<sup>-1</sup>) e 60 µL de solução II. A placa foi selada, invertida (20x) para homogeneização. Após um rápido spin realizou-se a adição de 60 µL de solução III em cada poço, a placa foi homogeneizada (inversão por 20x) e centrifugada como indicado acima. A placa foi incubada na estufa (90 °C por 30 min.), depois colocada no gelo por 10 min. e centrifugada novamente. O volume total do lisado foi transferido para placa de filtro (PALL) colocada sob a placa fundo V (Nunc) e centrifugada. A placa filtro foi descartada. Adicionamos 90 µL de isopropanol 100% em cada poço da placa de fundo V que contém o filtrado. A placa foi selada e homogeneizada por inversão (20x). Após homogeneização, a placa foi centrifugada por 45 minutos a 3220 g, para a precipitação do DNA plasmideal, o sobrenadante foi removido e foram adicionado 200 µL de etanol 70% gelado, seguido de uma centrifugação de 5 min.a 3220 g. O sobrenadante foi removido. A placa foi incubada em estufa em 37 °C por 15min para evaporação dos resíduos de etanol e o DNA foi ressuspendido em 20 µL de água Milli-Q.

Para a avaliação da integridade do DNA obtido, os plasmídios extraídos foram submetidos à eletroforese em gel de agarose e quantificados em espectrofotômetro (Nanodrop® Espectrophotometer ND-1000) como foi descrito previamente.

#### **4.2.7 Sequenciamento**

Para a busca de SNPs foram realizados dois tipos de sequenciamento:

1º Foi sequenciado 40 ng em 4 µL dos produtos da PCR obtidos a partir das ampliações de gDNA dos diferentes cultivares e espécies de café. Foram utilizados *primers* de GC que possuem os adaptadores M13For e M13Rev. Neste caso, foram seqüenciados 10 ng por µL, os mesmos *primers* foram usados durante as reações de seqüenciamento. 2º Para cada GC e para cada uma das amostras de gDNA de café, foram sequenciados 10 clones (400 ng de plasmídeo). O seqüenciamento dos clones foi realizado utilizando os *primers* universais *forward* e *reverse*, que se ligam em uma região dentro do vetor, mais próxima ao inserto clonado no vetor pCR®II.

Para o seqüenciamento, foi utilizado o método enzimático com base na síntese de DNA *in vitro*, na presença de nucleotídeos trifosfatados terminadores de cadeia (SANGER; NICKLEN; COULSON, 1977).

Todos os seqüenciamentos foram realizados com o kit “Big Dye Terminator” (Applied Biosystems) de acordo com o protocolo do fabricante. As reações de seqüenciamento foram realizadas no termociclador modelo PTC-100 (MJ Research, com o seguinte programa: Temperatura de desnaturação inicial de 96 °C, por 1min. seguidos de 35 ciclos de amplificação compostos de 3 etapas: 96 °C (desnaturação), por 10 seg.; 55 °C (anelamento), por 5 seg.; 60 °C (extensão), por 4 min. e 4 °C etapa final. Os produtos amplificados da reação de seqüenciamento foram precipitados com a adição de 40 µL de isopropanol 75% em cada cavidade da microplaca de PCR que continha as amostras. As placas contendo as amostras foram vedadas com adesivos e centrifugadas a 3220 g por 45 min. aos 20 °C. O isopropanol foi removido, a placa foi invertida sobre um papel toalha e realizou-se um rápido spin (1.000 g, 1 min.) para a retirada de todo isopropanol. Foi adicionado 200 µL de etanol 70% em cada cavidade da microplaca de PCR que continha as amostras, a placa vedada foi centrifugada por 10 min. aos 20 °C, o etanol foi removido, a placa foi invertida sobre um papel toalha e foi efetuado um rápido spin (1000 g, 1 min.) para a retirada de

todo etanol. As placas foram colocadas em estufa aos 37 °C por 30 min. para secagem.

Depois, foram adicionados 10 µL de Formamida Hi-Di (Applied Biosystems) em cada poço da microplaca, as reações foram submetidas ao termociclador modelo PTC-100 (MJ Research) para desnaturação e logo em seguida foram colocadas no seqüenciador ABI 3130xl Genetic Analyser (Applied Biosystems).

#### **4.2.8 Análise das seqüências**

As seqüências nucleotídicas obtidas foram analisadas a partir do uso do programa “Sequencing analysis” versão 5.2 da Applied Biosystems. Após analisar as seqüências, foi realizado o alinhamento das mesmas por meio do programa ClustalW (<http://clustalw.genome.ad.jp/>). Para traduzir as seqüências nucleotídicas encontradas foi utilizado o programa ExpASy (<http://expasy.org/tools/dna.html>).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Caracterização fisiológica

Nos anos de 2008 e 2009 durante o período de estresse, observamos que as medidas de potenciais hídricos ( $\Psi_{am}$ ) das plantas sem irrigação (NI) do cultivar Rubi sempre foram mais negativos que os potenciais hídricos das plantas sem irrigação (NI) do cultivar I59 (Gráfico 1 e 2) nos pontos de colheita P2 e P5. Essas medidas também mostram que o estresse hídrico do ano 2008 foi mais severo que o estresse hídrico do ano 2009, com valores de potenciais ao dia mais baixos para os dois cultivares no primeiro ano quando comparado ao segundo ano. Os dados obtidos para os dois pontos de colheita nos anos de 2008 e 2009 sugerem que o cultivar I59 é mais tolerante ao estresse hídrico quando comparado ao cultivar Rubi.

De acordo com Pinheiro (2005), os mecanismos fisiológicos associados à tolerância da espécie de *C. canephora* cv. Conilon parecem estar fortemente relacionados à grande sensibilidade estomática induzida pelo déficit hídrico no solo, aspecto este diretamente relacionado aos centros de origem das espécies de café, situados nas florestas tropicais da África. *C. arabica* é nativo dos sub-bosques das florestas de terras altas da Etiópia (1600-2800 m de altitude), onde as temperaturas apresentam pouca variação sazonal (15-20°C) (SILVA; MAZZAFERA, 2008). Estudos têm confirmado que plantas adaptadas à seca são caracterizadas por um sistema radicular profundo e vigoroso (PINHEIRO, 2005). De acordo com Taiz e Zeiger (2004), sob condições normais de crescimento a parte aérea funciona como o principal dreno de fotoassimilados; contudo, quando a planta está submetida ao déficit hídrico, algumas plantas desenvolvem raízes mais profundas que passam a receber uma maior proporção de assimilados, uma vez que a demanda energética da parte aérea é diminuída

devido à inibição da expansão foliar. Sob estresse moderado, a inversão na relação fonte-dreno favorece o crescimento do sistema radicular que, desta maneira, aprofunda suas raízes rumo às camadas mais inferiores e úmidas do solo. Comparando o sistema radicular dos cultivares Rubi e I59 das plantas analisadas com as medidas de potenciais, Marraccini et al. (2009) observaram uma maior biomassa seca e profundidade dos pivôs para o cultivar I59 do que para o cultivar Rubi. Com as medidas de potenciais, esses resultados constituem argumentos para afirmar que o cultivar I59 é mais tolerante à seca que o cultivar Rubi.

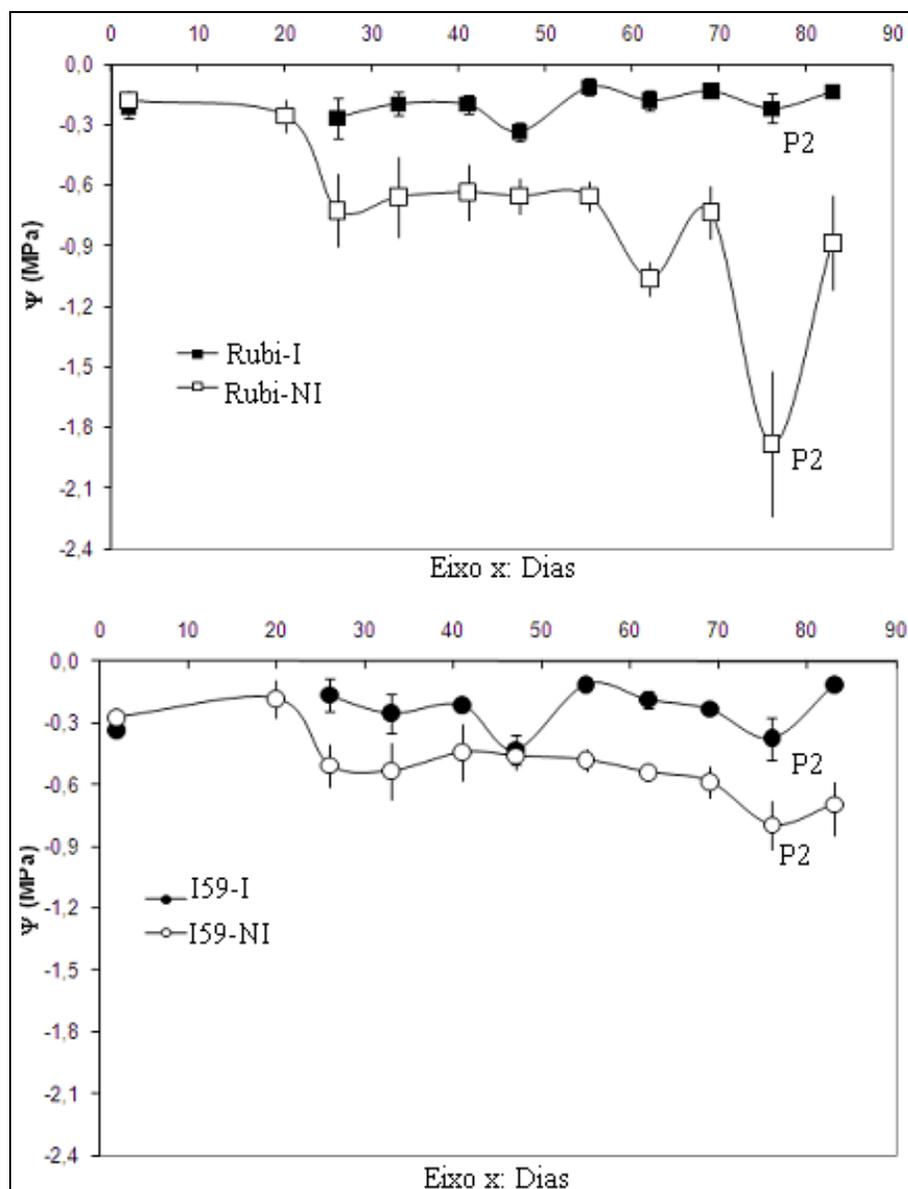


Gráfico 1 Evoluções dos potenciais de base ( $\Psi_{am}$ ) em folhas durante o estabelecimento do stress hídrico no ano 2008 (30/06/08 até o 19/09/08). A comparação é feita entre os cultivares Rubi (acima) e I59 (abaixo) irrigados (I) e não irrigado (NI). A abscissa x é referente aos dias quando o  $\Psi_{am}$  é expresso em MPa (Mega Pascal)

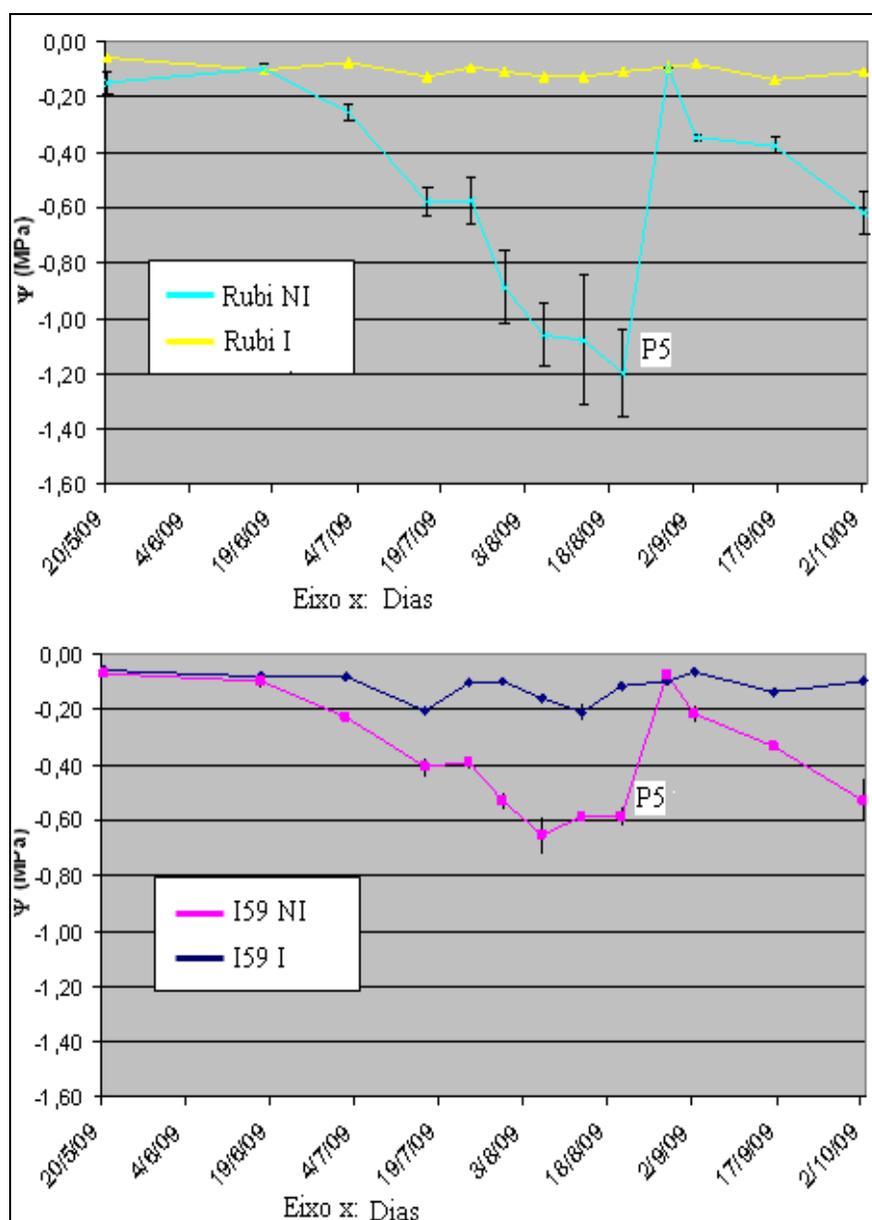


Gráfico 2 Evoluções dos potenciais de base ( $\Psi_{am}$ ) em folhas de *C. arabica* durante o estabelecimento do stress hídrico no ano 2009 (20/05/09 até o 02/10/09). A comparação é feita entre os cultivares Rubi (acima) e I59 (abaixo) irrigados (I) e não irrigado (NI). A abscissa é referente aos dias quando o  $\Psi_{am}$  é expresso em MPa (Mega Pascal)

### 5.2.1 Expressão do gene *NAC-RD26*

Para o ano 2008, os níveis de expressão do gene *NAC-RD26* (Contig 12424) nos dois cultivares apareceram como relativamente baixos e iguais nos pontos P1 e P2-I (sem estresse hídrico). Para os dois cultivares, foi observado um aumento da expressão desse gene com estresse hídrico nas folhas das plantas não irrigadas durante o período de seca (P2-NI) (Gráfico 3A). Nesse caso, esse aumento nas folhas de P2-NI quando comparado com as folhas de P2-I pareceu ser maior no cultivar Rubi (6,4 x) quando comparado ao cultivar I59 (3,3x). Ao contrário, no ano 2009, não se observou aumentos da expressão do gene *NAC-RD26* com o estresse hídrico para os dois cultivares durante o período de seca (P5) (Gráfico 3B). Observou-se uma leve queda de expressão para esse gene nas folhas do cultivar I59 P5-NI em comparação com as folhas irrigadas, enquanto os níveis de expressão desse gene foram similares ao observado na folhas do cultivar Rubi P5.

### 5.2.2 Expressão do gene *RD29*

Como observado para o gene *NAC-RD26*, os níveis de expressão do gene *RD29* (Contig 5590) apareceram baixos sem estresse hídrico (P1 e P2-I), e aumentaram com o estresse hídrico nas folhas das plantas não irrigadas durante o período de seca (P2-NI) (Gráfico 3C). Nesse caso, esse aumento foi muito maior no cultivar Rubi (10,6x) quando comparado ao cultivar I59 (2x). Uma situação diferente foi observada para a expressão desse gene no ano 2009, mostrando uma baixa expressão independentemente dos pontos de colheita (P4 e P5) e dos tratamentos (I e NI) (Gráfico 3D).

### 5.2.3 Expressão do gene mannose-6-fosfato redutase (*M6FR*)

No ano 2008, para os dois cultivares, os níveis de expressão do gene *M6FR* (Contig 13762) nas folhas sem estresse hídrico parecem maiores no ponto P2-I, quando comparado ao ponto controle P1 (Gráfico 3E). Com o estresse hídrico, é possível verificar um aumento da expressão desse gene (5,5x) para os dois cultivares. Durante o ponto P2 e independentemente das condições de irrigação, a expressão do gene *M6FR* pareceu ser sempre maior (2x) nas folhas do cultivar I59 que nas folhas do Rubi.

Durante os pontos de colheita P4 e P5 do ano 2009, todos os níveis de expressão do gene *M6FR* apareceram mais baixos do que no ano anterior (Gráfico 3F). Mesmo assim, foi possível observar um aumento da expressão desse gene com o estresse hídrico maior nas folhas do I59 do que nas folhas de Rubi.

### 5.2.4 Expressão do gene *GCI0*

Para o gene *GCI0* (Contig13580), não foram observadas variações da expressão nas folhas do cultivar I59 com as diferentes condições de irrigação nos pontos de colheita P1 e P2 do ano 2008. Entretanto, a expressão desse gene nas folhas do cultivar Rubi foi maior que no I59, aumentando (3x) durante o período de seca sem irrigação (P2-NI) (Gráfico 3G).

Durante o ano 2009, foi possível detectar a expressão do gene *GCI0* nas folhas do cultivar I59 nos dois pontos de colheita (P4 e P5). Em comparação com o ano de 2008, os níveis de expressão são maiores para os dois cultivares no ano de 2009 (Gráfico 3H). Em relação ao nível de expressão nas folhas não estressadas (P5-I), é possível verificar um aumento da expressão do gene *GCI0*

com o estresse hídrico (P5-NI) maior (30x) nas folhas do Rubi do que nas folhas do I59 (3x).

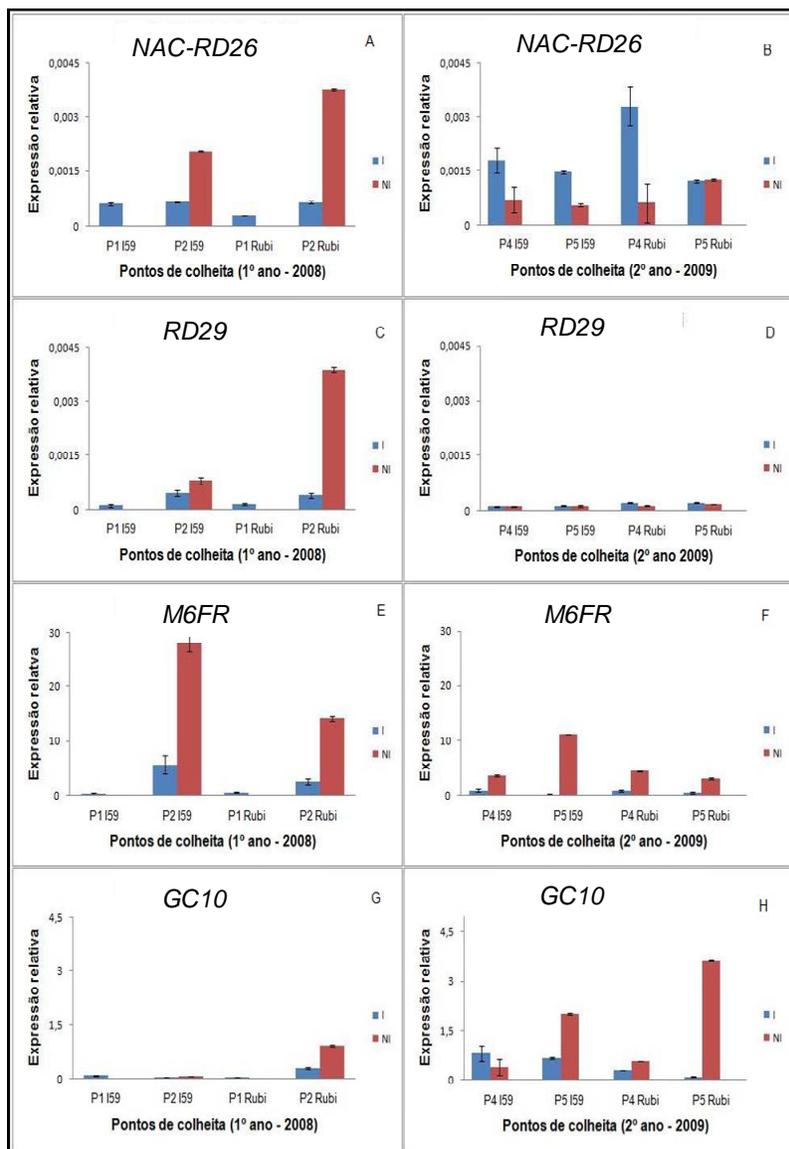


Gráfico 3 Perfis de expressão dos genes candidatos *NAC-RD26* (A, B), *RD29* (C, D), *M6FR* (E, F) e *GC10* (G, H) nas folhas das plantas colhidas durante os anos de 2008 (esquerda) e de 2009 (direita). Cultivares (I59: Iapar59 e Rubi), pontos de colheita (P) e os tratamentos (I: irrigado e NI: não irrigado) são indicados nas figuras

### **5.2.5 Expressão dos genes *AC* e *OEC***

É possível verificar que os níveis de expressão foram praticamente inalterados com o estresse hídrico do ano 2008 para cultivar I59, e para o cultivar Rubi eles baixaram (Gráfico 4A e C).

Em 2009, observou-se para os dois genes níveis de expressão altos na condição I (P5) no cultivar I59, enquanto os níveis foram praticamente iguais nas folhas irrigadas (I) e não irrigadas (NI) do cultivar Rubi (P5) (Gráfico 4B e D).

### **5.2.6 Expressão do gene *DREB***

No ano 2008 no ponto de controle P1 (sem estresse hídrico), a expressão do gene *DREB* (Contig 14421) foi maior (10x) nas folhas do cultivar I59 quando comparado as folhas do Rubi (Gráfico 4E). No ponto P2, para os dois cultivares, os níveis de expressão do gene *DREB* apareceram iguais nas folhas irrigadas e mostraram uma queda com o estresse hídrico.

Em 2009 os níveis de expressão parecem muito mais baixos do que em 2008 (Gráfico 4F). Isso é observado em particular nas folhas do cultivar I59 nos pontos P4 e P5. No ponto P5, o cultivar Rubi apresentou maiores níveis de expressão. Esses níveis não foram alterados com o nível de estresse aplicado as plantas.

### **5.2.7 Expressão do gene *CCoAMT***

No ano 2008, a expressão do gene *CCoAMT* (Contig 6147) foi igual para os dois cultivares no ponto controle P1. No ponto P2-I, a expressão desse gene foi maior (3x) nas folhas do I59 quando comparados as folhas do cultivar

Rubi. Para os dois cultivares, a expressão baixou com o estresse hídrico (P2-NI), mas ficou sempre maior (4,5x) no I59 quando comparado ao Rubi (Gráfico 4G).

Como em 2008, a expressão do gene *CCoAMT* apareceu maior (7x) nas folhas irrigadas do I59 do que nas folhas irrigadas do Rubi no ponto P5. Foi possível observar uma queda de expressão nas folhas do cultivar I59 submetidas ao estresse hídrico (Gráfico 4H).

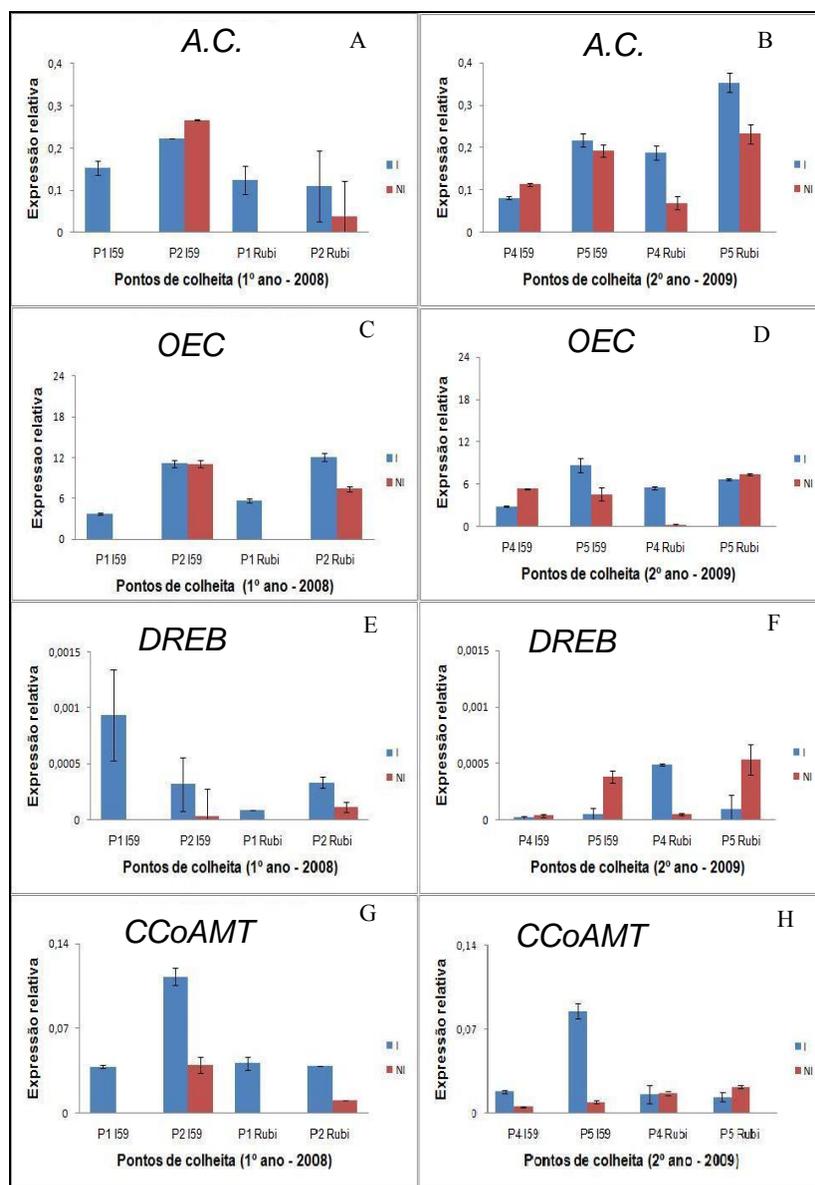


Gráfico 4 Perfis de expressão dos genes candidatos *AC* (A, B), *OEC* (C, D), *DREB* (E, F) e *CCoAMT* (G, H) nas folhas colhidas durante os anos e 2008 (esquerda) e de 2009 (direita). Cultivares (I59: Iapar59 e Rubi), pontos de colheita (P) e os tratamentos (I: irrigado e NI: não irrigado) são indicados nas figuras

### 5.2.8 Expressão do gene *MYB61*

Para os dois anos de análise e para cada ponto de colheita, os níveis da expressão do gene *MYB61* (Contig 1751) foram sempre maiores nas folhas irrigadas do cultivar I59 quando comparadas as folhas do cultivar Rubi (Gráfico 5A e B). No ano 2008, as variações de expressão em relação ao estresse hídrico (P2I vs. NI) não foi significativo para os dois cultivares.

Em 2009, foi possível observar o efeito da seca sobre a expressão do gene *MYB61* com a queda de expressão desse gene nos dois cultivares no ponto P5. Nesse caso, e independentemente das condições hídricas, os níveis de expressão do gene *MYB61* sempre ficaram maiores nas folhas do I59 quando comparadas com as folhas do cultivar Rubi, única exceção no ponto P5 onde as duas condições permaneceram praticamente iguais.

### 5.2.9 Expressão do gene *RBCS*

Para se analisar a expressão do gene *RBCS* (Contig 9250), foram usados dois pares de *primers* específicos para diferentes isoformas, *RBCS-1A* e *RBCS-1B* respectivamente (Gráfico 5C a F).

No ano 2008, as expressões das isoformas *RBCS-1A* e *RBCS-1B* foram altas nas folhas do cultivar I59, mas não foram detectadas no cultivar Rubi. Nesse caso, as expressões das duas isoformas foram sempre maiores no ponto P2 do que no ponto P1 e não pareceram ser alteradas com o estresse hídrico (Gráfico 5C e E).

Em 2009, os perfis de expressão das isoformas *RBCS-1A* e *RBCS-1B* foram praticamente iguais aos perfis observados no ano de 2008. Assim a expressão dessas isoformas foram altas no cultivar I59 no ponto P4 e P5 (Gráfico 5D e F), mas não foram detectadas no cultivar Rubi.

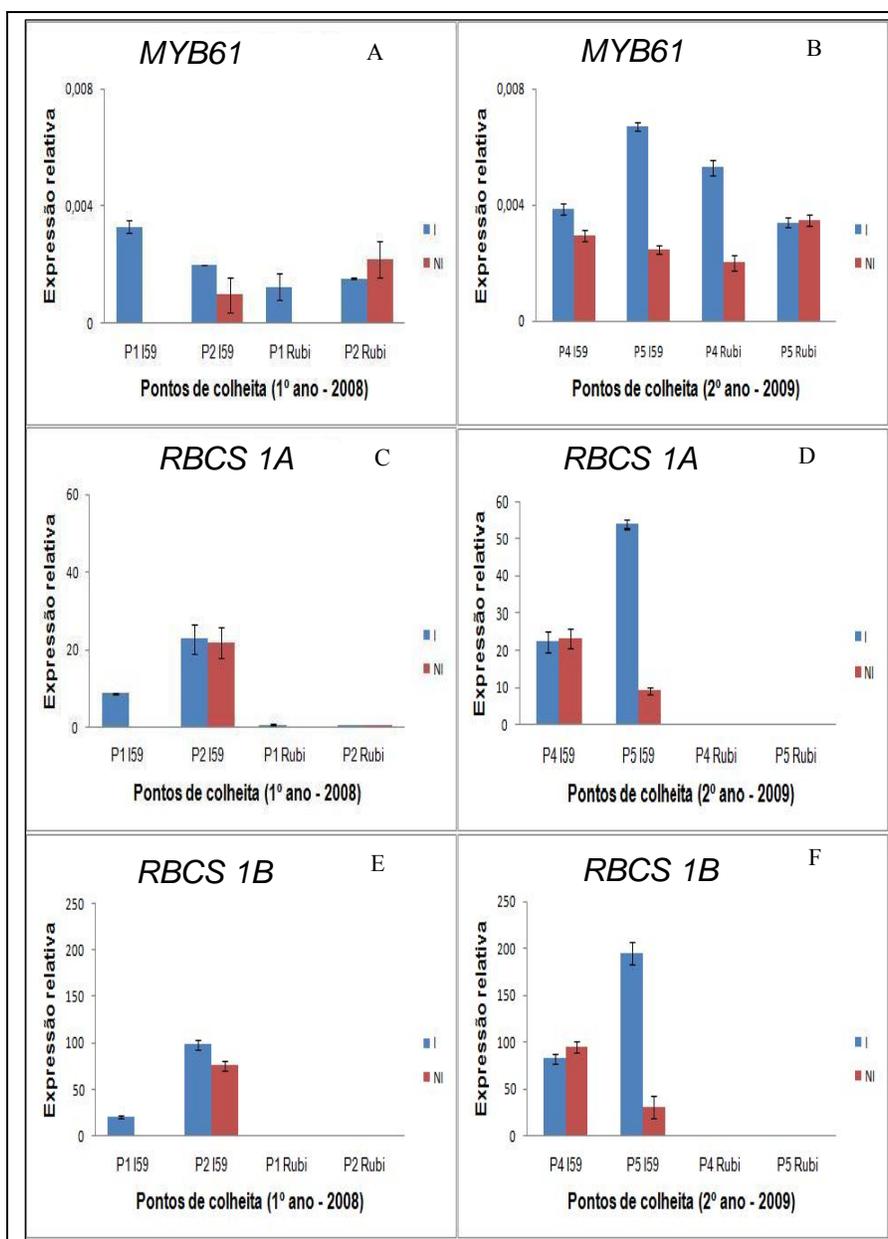


Gráfico 5 Perfis de expressão dos genes candidatos *MYB61* (A, B), *RBCS-1A* (C, D) e *RBCS-1B* (E, F) nas folhas colhidas durante os anos de 2008 (esquerda) e de 2009 (direita). Os cultivares (I59: Iapar59 e Rubi), pontos de colheita (P) e os tratamentos (I: irrigado e NI: não irrigado) são indicados nas figuras

### **5.2.10 Discussão: validação da expressão dos genes candidatos em cultivares de *C. arabica* cultivados no campo**

Nos resultados de qPCR existem vários fatores que interferem nas respostas de expressão dos GC como, por exemplo: o estresse hídrico, os dois anos de análise do experimento e os cultivares. Assim decidimos discutir os resultados em partes separadas.

#### **5.2.10.1 Efeito do estresse hídrico nos dois anos de estudo sobre a expressão dos GC**

As análises dos perfis de expressão dos GC mostram que o estresse hídrico afeta a expressão dos genes. Assim, foi possível verificar que a expressão dos genes *NAC-RD26* e *RD29* aumentaram com o estresse hídrico durante o ano 2008. Esse aumento foi observado durante os dois anos do ensaio para o gene *M6FR* (Gráfico 3E e F). Aumentos de expressão também foram observados durante o ano 2009 para o gene *GC10* (Gráfico 3G e H) nos dois cultivares e somente nas folhas de I59 para os genes *AC* e *OEC* no ponto P4 (Gráfico 4A à D). Para esses dois últimos genes, é interessante observar que os perfis de expressão são semelhantes para os dois anos do experimento. Isso pode estar relacionado com o funcionamento da fotossíntese, pois as proteínas correspondentes são implicadas nesse processo.

Ao contrário dos genes citados acima, foi observado que nos genes *DREB* e *CCOAMT* (Gráfico 4E à H) ocorreu queda de expressão com o estresse hídrico para os dois cultivares durante o ano 2008. e para o gene *MYB61* durante o ano 2009. Para o gene *CCoAMT*, essa queda de expressão foi também observada no ano seguinte (2009), mas somente nas folhas do cultivar I59. As

diferenças do nível de expressão dos genes nos dois anos do experimento podem ser explicadas em parte com as variações do nível de estresse recebidas pelas plantas. Assim, as medidas dos potenciais hídricos mostram que o estresse hídrico foi maior no ponto P2 de 2008 que no ponto P5 de 2009. Sendo possível comentar os resultados de expressão do gene *M6FR* que mostra altos níveis de expressão em relação ao gene endógeno *G3PDH*. Dessa forma, os níveis de expressão desse gene apareceram maiores no ano 2008 que no ano de 2009, corroborando perfeitamente com a hipótese de influência direta do estresse hídrico sobre o nível de resposta desse gene.

Foram observadas também variações de expressão entre os dois anos do experimento para os genes *RD29* e *GC10*. Assim, os perfis de expressão observados no ano 2008 para o gene *RD29* nos dois cultivares, não foram observados no ano 2009, durante o qual a expressão desse gene quase não foi observada. Uma situação diferente foi observada para o gene *GC10*. Nesse caso, a expressão desse gene somente foi observada no ano 2008 no cultivar Rubi enquanto no ano 2009, a expressão foi observada nos dois cultivares com um aumento nas condições de estresse. Assim, esses resultados sugerem que o gene *RD29* somente se expressa em condições de estresse hídrico severo enquanto a expressão do gene *GC10* tem uma maior expressão em condições de estresse hídrico moderado, fato que poderá ser testado em condições controladas ou contrastante no campo.

#### **5.2.10.2 Comparação dos perfis de expressão dos GC entre os cultivares I59 e Rubi de *C. arábica***

Considerando o cultivar I59 de *C. arábica* mais tolerante à seca que o cultivar Rubi, foi possível identificar GC que apresentam diferentes perfis de expressão nos dois cultivares com as condições de estresse hídrico que foram

aplicadas as plantas. Assim, as diferenças dos níveis de expressão observadas nos dois cultivares para os genes *MYB61*, *NAC-RD26*, *RD29*, *DREB*, *M6FR* e *GC10* podem destacar esses cultivares em relação ao estresse hídrico. Assim, para o *GC10*, nenhuma expressão foi observada nas folhas do I59 no ano 2008 e a sua expressão em condição de estresse no ano 2009 ficou mais baixa no I59 do que no Rubi.

Os resultados apresentado indicam também que os valores das expressões relativas dos genes *M6FR* e *GC10* são mais altos do que os dos genes *MYB61*, *NAC-RD26* e *RD29* que codificam para fatores de transcrição, pois esses genes sempre apresentam baixos níveis de expressão em comparação aos níveis de expressão do gene endógeno de referência *G3PDH*. Nesse caso vale ressaltar que o uso dos genes *M6FR* e *GC10* nas reações de qPCR podem ser privilegiados em relação ao uso dos outros genes.

Como mencionado acima, mesmo codificando para proteínas que assumem funções diferentes na fotossíntese, os perfis de expressão dos genes *AC* e *OEC* foram similares durante os dois anos de estudo para os dois cultivares. Os perfis mostram também respostas diferenciadas entre os dois cultivares sob condição de estresse hídrico. Assim no ano de 2008, a expressão desses genes não foi alterada no ponto P2 para o cultivar I59, enquanto que para o cultivar Rubi houve uma queda da expressão. Em 2009, a expressão desses dois genes aumentou na condição não irrigado (P5) para o cultivar I59, entretanto permaneceu estável para o cultivar Rubi. Essas respostas dos genes *AC* e *OEC* em condição irrigada poderão acarretar diferenças de funcionamento da fotossíntese entre os dois cultivares.

### **5.2.10.3 Comparação dos perfis de expressão dos GC em *C. arabica* e em *C. canephora***

Com exceção das variações de expressão dos genes que foram observadas nos dois anos do experimento e previamente discutidas, foi possível confirmar em *C. arabica*, os resultados de expressão de alguns GC obtidos em *C. canephora* (Figura 1). Assim, foram observados aumentos de expressão com o estresse hídrico para os genes *NAC-RD26* (2008), *RD29* (2008), *M6FR* (2008 e 2009) e *GC10* (2009) nas folhas de *C. arabica* como também descrito em *C. canephora*. Nesses casos, é interessante ressaltar que os genes *RD29* e *GC10* apresentaram maior expressão com estresse hídrico no cultivar Rubi de *C. arabica* e no clone 22 de *C. canephora*, ambos considerados sensíveis à seca, do que no cultivar I59 de *C. arabica* e no clone 14 de *C. canephora*, os quais são considerados tolerantes à seca. Pelo contrário, o gene *M6FR* mostrou uma expressão maior com estresse hídrico no cultivar I59 de *C. arabica* e no clone 14 de *C. canephora*, do que no cultivar Rubi de *C. arabica* e no clone 22 de *C. canephora*. Como o gene *M6FR* codifica para uma enzima que controla a síntese de manitol, a sua maior expressão em condição de estresse hídrico pode estar diretamente ligada à maior tolerância dessas plantas ao estresse hídrico.

Para alguns genes, não foi possível confirmar em *C. arabica* os resultados de expressão obtidos em *C. canephora*. Assim, quando foi observada uma queda de expressão com o estresse hídrico nos genes *AC* e *OEC* em *C. canephora*, no ano de 2009 em *C. arabica* os resultados mostram para o cultivar I59 com estresse hídrico aumento da expressão nos dois genes. Entretanto não foi verificado nenhum efeito no cultivar Rubi. Para os genes *CCoAMT* e *MYB61*, os perfis de expressão em *C. canephora* sempre foram maiores em folhas do clone 22 sensível à seca do que no clone 14, sem que fossem observados efeitos do estresse hídrico sobre a expressão desses genes. Em *C. arabica*, a expressão dos genes *CCoAMT* e *MYB61* foi maior nas folhas do cultivar I59 quando comparada a expressão dos genes nas folhas do cultivar Rubi e foi reduzida em condição de estresse hídrico particularmente no I59.

Em *C. canephora*, a expressão do gene *DREB* sempre foi maior no clone 14 em condição de estresse hídrico que no clone 22. Os resultados obtidos para os dois cultivares em *C. arabica*, mostram uma queda de expressão com o estresse hídrico no ano 2008, e uma maior expressão desses genes nas folhas I e NI de Rubi no P5 do ano 2009 do que no I59.

#### **5.2.10.4 Caso particular do gene *RBCS***

Os pares de *primers* *RBCS-1A* (B18224) e *RBCS-1B* (C18244), Alves et al. (2009) mostram uma queda da expressão dessas duas isoformas com o estresse hídrico nas folhas dos clones 14 e 22 de *C. canephora*. Os mesmos autores observaram também que a isoforma *RBCS-1B* se expressava em folhas de plantas adultas do cultivar I59, com uma leve queda de expressão em condição de estresse hídrico, mas não se expressava em folhas de plantas adultas do cultivar Rubi.

Os resultados obtidos nesse trabalho confirmaram a ausência de expressão da isoforma *RBCS-1B* em folhas de plantas jovens de Rubi e mostram também uma leve queda de expressão desse gene com o estresse hídrico no ano 2008, o mesmo foi observado no ano de 2009 nas folhas de plantas jovens de I59. A isoforma *RBCS-1A* apresentou perfil de expressão similar a isoforma *RBCS-1B* no ano de 2008 com a ausência de expressão na folhas de Rubi e uma alta expressão durante o período de seca do mesmo (P2), portanto sem apresentar queda significativa de expressão com o estresse hídrico. A expressão da isoforma *RBCS-1A* e *RBCS-1B* igualmente ao ano de 2008 foram detectadas para os pontos de colheita P4 e P5 no somente no cultivar I59 no ano 2009. Assim, nenhuma expressão das isoformas foram observadas no cultivar Rubi. Os resultados obtidos confirmaram que os pares de *primers* usados são específicos de isoformas diferentes do gene *RBCS* e que apresentam perfis de expressão

diferenciais em função dos cultivares de *C. arabica* analisados. O cultivar I59 (chamado de “sarchimor”) é oriundo de um cruzamento entre o híbrido de Timor 832/2 e o *C. arabica* Villa Sarchi 971/10 (SERÁ, 2001) enquanto o Rubi é considerado “100% Arabica”, não possui introgressão recente de *C. canephora* (CARVALHO, 2008). O fato das isoformas *RBCS-IA* e *RBCS-IB* não se expressar no Rubi enquanto se expressam no I59 sugere que essas isoformas correspondem a formas alélicas provavelmente oriundas de *C. canephora* por meio da introgressão ocorrida e oriunda do híbrido Timor 832/2. Trabalhos estão justamente em andamento no Laboratório de Genética Molecular da Embrapa Recursos Genéticos e Biotecnologia para se avaliar essa hipótese.

### 5.3 Análise dos polimorfismos (SNPs) do gene *DREB*

A partir da lista dos 28 gDNA avaliados, conseguimos as seqüências para 25 genótipos. Não foi possível analisar os genótipos Bourbon, E464, E238, E123B e *C. eugenioides*, pois o sequenciamento dos mesmos não permitiu a leitura das seqüências. Usando o gene *DREB* (Contig 14421) como seqüência de referência, as análises das seqüências obtidas mostram que (i) esse gene não contém introns e (ii) que não foram encontradas modificações de seqüências do tipo inserção/deleção na região codante do gene.

A busca de SNPs dentre todos os genótipos utilizados permitiu encontrar ao todo sete polimorfismos a partir das análises detalhadas dos eletroferogramas. Os polimorfismos 1, 2, 3 e 4 foram encontrados na região codante (Tabela 5A) e os polimorfismos 5, 6 e 7 estão localizados na região 3' UTR (região não codante) do gene *DREB* (Tabelas 5A e 5B).

Foi encontrada uma inserção de base no polimorfismo de número 6 onde é possível verificar a ausência (-) de um nucleotídeo G. Nos polimorfismos 1, 2, e 4, foram encontrados SNPs de transição (T>C), (C>T) e (G>A)

respectivamente. Entretanto polimorfismo 5 (T>C) e 7 (T>A) são chamados de transversão.

Os polimorfismos SNP<sub>DREB1</sub> e SNP<sub>DREB3</sub> alteram o aminoácido da proteína encontrada. O primeiro altera a segunda base do códon nº6 da proteína traduzida a partir do Contig 14421, trocando-se o aminoácido I (Isoleucina) para T (Treonina). O polimorfismo SNP<sub>DREB3</sub> altera a segunda base do códon número 213 trocando o aminoácido A (Alanina) para G (Glicina). Entretanto o SNP<sub>DREB2</sub> altera a terceira base do códon número 22, não modificando o aminoácido D (Aspartato), assim como acontece com o polimorfismo SNP<sub>DREB4</sub> que altera a terceira base do códon número 232, mas não modifica o aminoácido S (Serina) (Tabela 6).

Tabela 5A SNPs do gene *DREB* localizados na sequência codante. A sequência de referência do gene *DREB* (Contig 14421) é apresentada na tabela logo acima dos genótipos. Para cada um dos genótipos testados, os nucleotídeos divergentes (SNPs) da sequência de referência são apresentados em vermelho. Nucleotídeos iguais ao da sequência de referência são apresentados em amarelo. As posições dos SNPs na sequência de referência são indicadas na última linha da tabela. Genótipos não analisados por problemas nas sequências estão indicados por (R = refazer)

	SNP <sub>DREB1</sub> (T/C)	SNP <sub>DREB2</sub> (C/T)	SNP <sub>DREB3</sub> (C/G)	SNP <sub>DREB4</sub> (G/A)
<b>Gene <i>DREB</i></b>	<b>AATAT</b>	<b>GACTC</b>	<b>TGCAG</b>	<b>TCGAT</b>
<b>Genótipos</b>				
01-M. Novo	AATAT AACAT	GACTC GATTC	TGCAG TGGAG	TCGAT TCAAT
02-Catuai 25	AACAT	GACTC GATTC	TGCAG TGGAG	TCGAT TCAAT
03-Typica	AATAT AACAT	GACTC	TGCAG TGGAG	TCGAT TCAAT
04-Bourb.	R	R	R	R
05-Rubi	AATAT AACAT	GACTC GATTC	TGCAG TGGAG	TCGAT TCAAT
06-159	AATAT AACAT	GACTC GATTC	TGCAG TGGAG	TCGAT TCAAT

“Tabela 5A, continua”

“Tabela 5A, conclusão”				
07-E516	AA <b>T</b> AT AA <b>C</b> AT	GA <b>C</b> TC GA <b>T</b> TC	R	R
08-E464	R	R	R	R
09-E007	AA <b>T</b> AT AA <b>C</b> AT	GA <b>C</b> TC GA <b>T</b> TC	TGCAG TGCAG	TCGAT TC <b>A</b> AT
10-G237	AA <b>T</b> AT AA <b>C</b> AT	GA <b>C</b> TC GA <b>T</b> TC	TGCAG TGCAG	TCGAT TC <b>A</b> AT
11-E017	AA <b>C</b> AT AA <b>T</b> AT	GA <b>C</b> TC GA <b>T</b> TC	TGCAG TGCAG	TCGAT TC <b>A</b> AT
12-E238	R	R	R	R
13-E123B	R	R	R	TC <b>A</b> AT TCGAT
14-E123A	R	GA <b>C</b> TC GA <b>T</b> TC	TGCAG TGCAG	TCGAT TC <b>A</b> AT
15-UW099	AA <b>C</b> AT	GA <b>C</b> TC	TGCAG	TC <b>A</b> AT
16-UW002	AA <b>C</b> AT	GA <b>C</b> TC	TGCAG	TC <b>A</b> AT
18-C3001	AA <b>C</b> AT	GA <b>C</b> TC	TGCAG	TC <b>A</b> AT
19-Caneph.	R	GA <b>C</b> TC	R	R
20-G2011	AA <b>C</b> AT	GA <b>C</b> TC	TGCAG TGCAG	TC <b>A</b> AT
21-CI007	AA <b>C</b> AT	GA <b>C</b> TC	TGCAG	TC <b>A</b> AT
22-C4001	R	R	TGCAG TGCAG	TCGAT TC <b>A</b> AT
23-C.14	AA <b>C</b> AT	GA <b>C</b> TC	R	R
24-C.22	AA <b>T</b> AT AA <b>C</b> AT	GA <b>C</b> TC	TGCAG TGCAG	TC <b>A</b> AT
25-G2020	AA <b>C</b> AT	GA <b>C</b> TC	TGCAG	TC <b>A</b> AT
26-Psil.	AA <b>T</b> AT AA <b>C</b> AT	GA <b>C</b> TC GA <b>T</b> TC	TGCAG TGCAG	TCGAT TC <b>A</b> AT
27-Racem.	AA <b>T</b> AT	GA <b>T</b> TC	TGCAG	TC <b>A</b> AT
28- Eug.	R	R	R	
<b>Posição SNP</b>	<b>165</b>	<b>214</b>	<b>786</b>	<b>844</b>

Tabela 5B SNP do gene *DREB* localizados na sequência não codante 3'UTR. A sequência de referência do gene *DREB* (Contig 14421) é apresentada na tabela logo acima dos genótipos. Para cada um dos genótipos testados, os nucleotídeos divergentes (SNPs) da sequência de referência são apresentados em vermelho. Aminoácidos iguais ao da sequência de referência são apresentados em amarelo. As posições dos SNPs na sequência de referência são indicadas na última linha da tabela. Genótipos não analisados por problemas nas sequências estão indicados por (R = refazer)

Gene <i>DREB</i>	SNP <sub>DREB5</sub> (T/C)	SNP <sub>DREB6</sub> (-/C)	SNP <sub>DREB7</sub> (T/A)
<b>Genótipos</b>	<b>TG<b>T</b>AG</b>	<b>AG<b>T</b>A</b>	<b>AG<b>T</b>AC</b>
01-M. Novo	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
02-Catuai 25	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
03-Typica	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA GA <b>T</b> TC	AG <b>T</b> AC AG <b>A</b> AC
04-Bourbon	R	R	R
05-Rubi	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TAG	AG <b>T</b> AC
06-I59	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC
07-E516	R	R	R
08-E464	R	R	R
09-E007	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
10-G237	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
11-E017	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
12-E238	R	R	R
13-E123B	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
14-E123A	R	R	R
15-UW099	TG <b>C</b> AG	AG <b>T</b> TA	R-
16-UW002	TG <b>T</b> AG	AG <b>T</b> TA AG <b>T</b> TA	AG <b>T</b> AC AG <b>T</b> AC
18-C3001	TG <b>T</b> AG	AG <b>T</b> TA	AG <b>T</b> AC
19-Caneph.	R	R	R
20-G2011	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC
21-CI007	R	AG <b>T</b> TA	AG <b>T</b> AC
22-C4001	TG <b>C</b> AG TG <b>T</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
23-C.14	R	R	R
24-C.22	TG <b>C</b> AG TG <b>T</b> AG	AG <b>T</b> TA	AG <b>T</b> AC
25-G2020	TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC
26-Psil.	TG <b>T</b> AG TG <b>C</b> AG	AG <b>T</b> TA	AG <b>T</b> AC AG <b>A</b> AC
27-Racem.	TG <b>T</b> AG	AG <b>T</b> TA	AG <b>T</b> AC
28-Eug.	R	R	R
<b>Posição SNP</b>	<b>879</b>	<b>746</b>	<b>774</b>

Tabela 6 Sequências de aminoácido deduzidas das regiões nucleotídicas contendo os SNP, localizados na parte codante da proteína DREB. Para cada um dos genótipos testados, os aminoácidos divergentes da proteína de referência são apresentados em vermelho. Aminoácidos iguais aos da proteína de referência são apresentados em amarelo. As posições dos aminoácidos referentes aos SNPs na sequência da proteína de referência são indicadas na última linha da tabela. Genótipos não analisados por problemas nas sequências estão indicados por (R = refazer)

Proteína DREB	SNP <sub>DREB1</sub>	SNP <sub>DREB2</sub>	SNP <sub>DREB3</sub>	SNP <sub>DREB4</sub>
Genótipos	FSICY	FPDSS	QCAVG	WSYSI
01-M. Novo	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
02- Catuai 25	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
03-Typica	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
04-Bourb.	R	R	R	R
05-Rubi	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
06-I59	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
07-E516	FSICY FSICY	FPDSS	QCAVG	WSYSI
08-E464	R	R	R	R
09-E007	FSICY FSICY	FPDSS	QCGVG	WSYSI
10-G237	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
11-E017	FSICY FSICY	FPDSS	QCAVG	WSYSI
12-E238	R	R	R	R
13-E123B	FSICY FSICY	R	R	WSYSI
14-E123A	R	FPDSS	QCAVG QCGVG	WSYSI
15-UW099	FSICY	FPDSS	QCAVG	WSYSI
16-UW002	FSICY	FPDSS	QCAVG	WSYSI
18-C3001	FSICY	FPDSS	QCGVG	WSYSI
19-Canep.	FSICY	FPDSS	R	R
20-G2011	FSICY	FPDSS	QCAVG QCGVG	WSYSI
21-CI007	FSICY FSICY	FPDSS	QCGVG	WSYSI
22-C4001	FSICY	FPDSS	QCAVG QCGVG	WSYSI
23-C.14	FSICY	FPDSS	R	R
24-C.22	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI
25-G2020	FSICY	FPDSS	QCAVG	WSYSI
26-Psila.	FSICY FSICY	FPDSS	QCAVG QCGVG	WSYSI

“Tabela 6, continua”

"Tabela 6, conclusão"				
27-Racem	FS CY	FPDSS	QC GVG	WSY SI
28-Eug.	R	R	R	R
Posição aa	6	22	213	232

### 5.3.1 Análise dos polimorfismos (SNPs) do gene *NAC-RD26*

Em relação ao gene *NAC-RD26* (Conter 12424), dos 28 gDNA amplificados, conseguimos sequências para 24 genótipos. Assim, não foi possível analisar os genótipos Catuaí, E123A, Canephora G21 (SG1) e o clone 22 de Conilon, porque os seqüenciamentos dos mesmos não ficaram bons. A partir da análise detalhada dos eletroferogramas das sequências dos 24 genótipos, não foram encontrados introns e modificações de sequências do tipo inserção/deleção na região codante do gene. A busca de SNPs revelou a presença de 7 polimorfismos (Tabela 7A e 7B), todos eles localizados na região codante do gene *NAC-RD26*. O polimorfismo 1 se caracteriza por uma transversão no SNP que troca o nucleotídeo G pelo C. Para os polimorfismos 2 (A>G), 3 (C>T), 4 (T>G), 5 (G>A), 6 (C>T) e 7, (G>A) todos são de SNP do tipo transição.

Os polimorfismos  $SNP_{NAC1}$ ,  $SNP_{NAC4}$  e  $SNP_{NAC5}$  alteram os aminoácidos da proteína deduzida do Contig 12424 (Tabelas 8A e 8B). O polimorfismo 1 altera a segunda base do códon 47 trocando o aminoácido G (Glicina) pelo aminoácido A (Alanina). O  $SNP_{NAC4}$  altera a terceira base do códon 309, trocando o aminoácido S (Serina) pelo aminoácido R (Arginina). O  $SNP_{NAC5}$  altera a segunda base do códon 323 trocando o aminoácido S (Serina) pelo aminoácido N (Asparagina). Para os polimorfismos 2, 3, 6 e 7, respectivamente referentes aos aminoácidos Glutamina ( $Q_{295}$ ), Serina ( $S_{305}$ ), Prolina ( $P_{324}$ ) e Treonina ( $T_{329}$ ), as modificações são silenciosas, alterando em todos os casos a terceira base dos códons sem modificar o aminoácido codificado.

Tabela 7A SNPs do gene *NAC-RD26* localizados na sequência codante. A sequência de referência do gene *NAC-RD26* (Contig 12424) é apresentada na tabela logo acima dos genótipos. Para cada um dos genótipos testados, os nucleotídeos divergentes (SNPs) da sequência de referência são apresentados em vermelho. Nucleotídeos iguais ao da sequência de referência são apresentados em amarelo. As posições dos SNPs na sequência de referência são indicadas na última linha da tabela. Genótipos não analisados por problemas nas sequências estão indicados por (R = refazer)

	SNP <sub>NAC1</sub> (G/C)	SNP <sub>NAC2</sub> (A/G)	SNP <sub>NAC3</sub> (C/T)	SNP <sub>NAC4</sub> (T/G)
<b>Gene <i>NAC-RD26</i></b>	<b>AGGAG</b>	<b>CAAGC</b>	<b>AGCGG</b>	<b>AGTCA</b>
<b>Genótipos</b>				
01-M.Novo	R	CAAGC	AGCGG	AGTCA
02-Catuai 25	R	CAAGC	AGTGG	AGTCA
03-Typica	AGGAG	R	R	R
	AGGAG			
04-Bourbon	R	CAAGC	AGCGG	AGTCA
		CAAGC	AGTGG	AGTCA
05-Rubi	R	CAAGC	AGCGG	AGTCA
		CAAGC	AGTGG	AGTCA
06-I59	R	CAAGC	AGCGG	AGTCA
		CAAGC	AGTGG	AGTCA
07-E516	AGGAG	CAAGC	AGCGG	AGTCA
08-E464	R	CAAGC	AGCGG	AGTCA
		CAAGC	AGTGG	AGTCA
09-E007	AGGAG	CAAGC	AGCGG	AGTCA
	AGGAG	CAAGC	AGTGG	AGTCA
10-G237	AGGAG	CAAGC	AGCGG	AGTCA
	AGGAG	CAAGC	AGTGG	AGTCA
11-E017	AGGAG	CAAGC	AGCGG	AGTCA
12-E238	R	CAAGC	AGCGG	AGTCA
		CAAGC	AGTGG	AGTCA
13-E123B	R	CAAGC	AGCGG	AGTCA
		CAAGC	AGTGG	AGTCA
14-E123A	R	R	R	R
15-UW099	AGGAG	CAAGC	AGCGG	AGTCA
	AGGAG	CAAGC	AGTGG	AGTCA
16-UW002	AGGAG	CAAGC	AGCGG	AGTCA
18-C3001	AGGAG	CAAGC	AGCGG	AGTCA
	AGGAG	CAAGC	AGTGG	AGTCA
19-Canep.	R	R	R	R
20-G2011	R	CAAGC	AGCGG	AGTCA
21-CI007	R	CAAGC	AGCGG	AGTCA
22-C4001	R	CAAGC	AGCGG	AGTCA
23-C.14	R	CAAGC	AGCGG	AGTCA
24-C.22	R	R	R	R
25-G2020	AGGAG	CAAGC	AGCGG	AGTCA
26-Psil.	R	CAAGC	AGTGG	AGTCA
27-Racem.	R	CAAGC	AGTGG	AGTCA
28-Eug.	AGGAG	CAAGC	AGTGG	AGTCA
<b>Posição SNP</b>	<b>276</b>	<b>1022</b>	<b>1051</b>	<b>1063</b>

Tabela 7B SNPs do gene *NAC-RD26* localizados na sequência codante. A sequência de referência do gene *NAC-RD26* (Contig 12424) é apresentada na tabela logo acima dos genótipos. Para cada um dos genótipos testados, os nucleotídeos divergentes (SNPs) da sequência de referência são apresentados em vermelho. Nucleotídeos iguais ao da sequência de referência são apresentados em amarelo. As posições dos SNPs na sequência de referência são indicadas na última linha da tabela. Genótipos não analisados por problemas nas sequências estão indicados por (R = refazer)

Gene <i>NAC-RD26</i>	SNP <sub>NAC5</sub> (G/A)	SNP <sub>NAC6</sub> (C/T)	SNP <sub>NAC7</sub> (G/A)
<b>Genótipos</b>	<b>CAGCC</b>	<b>CCCCT</b>	<b>ACGCA</b>
01-M. Novo	CAGCC CA <del>A</del> CC	CC <del>T</del> CT	AC <del>G</del> CA AC <del>A</del> CA
02- Catuai 25	R	R	R
03-Typica	R	R	R
04-Bourbon	CAGCC CA <del>A</del> CC	CC <del>T</del> CT CCCCT	AC <del>A</del> CA ACGCA
05-Rubi	CAGCC CA <del>A</del> CC	CC <del>T</del> CT	AC <del>G</del> CA AC <del>A</del> CA
06-I59	CAGCC CA <del>A</del> CC	CC <del>T</del> CT CCCCT	AC <del>G</del> CA AC <del>A</del> CA
07-E516	CAGCC	CCCCT	ACGCA
08-E464	CA <del>A</del> CC CAGCC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
09-E007	CAGCC CA <del>A</del> CC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
10-G237	CAGCC CA <del>A</del> CC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
11-E017	CAGCC	CCCCT	ACGCA
12-E238	CAGCC CA <del>A</del> CC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
13-E123B	CAGCC CA <del>A</del> CC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
14-E123A	R	R	R
15-UW099	CAGCC CA <del>A</del> CC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
16-UW002	CAGCC	CC <del>T</del> CT	ACGCA
18-C3001	CAGCC CA <del>A</del> CC	CCCCT CC <del>T</del> CT	ACGCA AC <del>A</del> CA
19-Canep.	R	R	R
20-G2011	CAGCC	CC <del>T</del> CT	ACGCA
21-CI007	CAGCC	CC <del>T</del> CT	ACGCA
22-C4001	CAGCC	CC <del>T</del> CT	ACGCA AC <del>A</del> CA
23-C.14	CAGCC	CC <del>T</del> CT	ACGCA
24-C.22	R	R	R
25-G2020	CAGCC	CCCCT	ACGCA
26-Psil.	CA <del>A</del> CC	CCCCT	AC <del>A</del> CA
27-Racem.	CA <del>A</del> CC	CC <del>T</del> CT	ACGCA
28-Eug.	CAGCC CA <del>A</del> CC	CC <del>T</del> CT	AC <del>G</del> CA AC <del>A</del> CA
<b>Posição SNP</b>	<b>1104</b>	<b>1108</b>	<b>1123</b>

Tabela 8A Sequências de aminoácido deduzidas das regiões nucleotídicas contendo os SNP localizados na parte codante da proteína NAC-RD26. Para cada um dos genótipos testados, os aminoácidos divergentes da proteína de referência são apresentados em vermelho. Aminoácidos iguais ao da sequência de referência são apresentados em amarelo. As posições dos aminoácidos referentes aos SNPs na sequência da proteína de referência são indicadas na última linha da tabela. Os genótipos não analisados por problemas nas sequências estão indicados por (R = refazer)

	SNP <sub>NAC1</sub>	SNP <sub>NAC2</sub>	SNP <sub>NAC3</sub>
<b>Proteína NAC-RD26</b>	<b>IIG<b>E</b>EI</b>	<b>FP<b>Q</b>AG</b>	<b>VQ<b>S</b>SGL</b>
<b>Genótipos</b>			
01-M. Novo	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
02-Catuai 25	R	R	R
03-Typica	II <b>A</b> EI IIG <b>E</b> EI	R	R
04-Bourbon	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
05-Rubi	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
06-I59	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
07-E516	IIG <b>E</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
08-E464	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
09-E007	IIG <b>E</b> EI II <b>A</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
10-G237	II <b>A</b> EI IIG <b>E</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
11-E017	IIG <b>E</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
12-E238	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
13-E123B	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
14-E123A	R	R	R
15-UW099	IIG <b>E</b> EI II <b>A</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
16-UW002	IIG <b>E</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
18-C3001	IIG <b>E</b> EI II <b>A</b> EI	FP <b>Q</b> AG	VQ <b>S</b> SGL
19-Canep.	R	R	R
20-G2011	R	FP <b>Q</b> AG	VQ <b>S</b> SGL
21-CI007	R	FP <b>Q</b> AG	VQ <b>S</b> SGL

“Tabela 8A, continua”

“Tabela 8A, conclusão”

22-C4001	R	FPQAG	VQSGL
23-C.14	R	FPQAG	VQSGL
24-C.22	R	R	R
25-G2020	IIGEI	FPQAG	VQSGL
26-Psil.	R	FPQAG	VQSGL
27-Racem.	R	FPQAG	VQSGL
28-Eug.	IIAEI	FPQAG	VQSGL
<b>Posição aa</b>	<b>47</b>	<b>295</b>	<b>305</b>

Tabela 8B Sequências de aminoácido deduzidas das regiões nucleotídicas contendo os SNP localizados na parte codante da proteína NAC-RD26. Para cada um dos genótipos testados, os aminoácidos divergentes da proteína de referência são apresentados em vermelho. Aminoácidos iguais ao da sequência de referência são apresentados em amarelo. As posições dos aminoácidos referentes aos SNPs na sequência da proteína de referência são indicadas na última linha da tabela. Os genótipos não analisados por problemas de seqüenciamentos são indicados (R = refazer)

	SNP <sub>NAC4</sub>	SNP <sub>NAC5</sub>	SNP <sub>NAC6</sub>	SNP <sub>NAC7</sub>
<b>Proteína NAC-RD26</b>	<b>LRSHR</b>	<b>NASPL</b>	<b>PHTHS</b>	
<b>Genótipos</b>				
01-M. Novo	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
02-Catuai 25	R	R	R	
03-Typica	R	R	R	
04-Bourbon	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
05-Rubi	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
06-I59	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
07-E516	LRSHR	NASPL	PH <del>T</del> HS	
08-E464	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
09-E007	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
10-G237	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
11-E017	LRSHR	NASPL	PH <del>T</del> HS	
12-E238	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
13-E123B	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS	
14-E123A	R	R	R	

“Tabela 8B, continua”

“ Tabela 8B, conclusão”

15-UW099	LRSHR LR <del>R</del> HR	NASPL NA <del>N</del> PL	PH <del>T</del> HS
16-UW002	LRSHR	NASPL	PH <del>T</del> HS
18-C3001	LRSHR LR <del>R</del> HR	NA <del>N</del> PL NA <del>N</del> PL	PH <del>T</del> HS
19-Canep.	R	R	R
20-G2011	LRSHR	NASPL	PH <del>T</del> HS
21-C1007	LRSHR	NASPL	PH <del>T</del> HS
22-C4001	LRSHR	NASPL	PH <del>T</del> HS
23-C.14	LRSHR	NASPL	PH <del>T</del> HS
24-C.22	R	R	R
25-G2020	LRSHR	NASPL	PH <del>T</del> HS
26-Psil.	LRSHR	NA <del>N</del> PL	PH <del>T</del> HS
27-Racem.	LRSHR	NA <del>N</del> PL	PH <del>T</del> HS PH <del>T</del> HS
28-Eug.	LR <del>R</del> HR	NA <del>N</del> PL	PH <del>T</del> HS
<b>Posição aa</b>	<b>47</b>	<b>323/324</b>	<b>329</b>

### 5.3.2 Análise dos polimorfismos (SNPs) do gene *RD29*

Em relação ao gene *RD29* (Contig 5590), e dentro da lista de 28 gDNA, não foram obtidas sequências para 11 genótipos testados. Da análise detalhada dos eletroferogramas, não foram encontrados introns e modificações de sequências do tipo inserção/deleção na região codante do gene analisado.

A busca de SNPs não revelou a existência de polimorfismos nas regiões 5'- e 3'-UTR. Assim, os 6 polimorfismos identificados (Tabelas 9A e B) são todos localizados na região codante do gene *RD29*, no último quarto (região C-terminal) da proteína deduzida do Contig 5590. Cinco SNPs são de “transição” ( $\text{SNP}_{RD29-2}$ ,  $\text{SNP}_{RD29-4}$  e  $\text{SNP}_{RD29-5}$  G>A,  $\text{SNP}_{RD29-3}$  A>G e  $\text{SNP}_{RD29-6}$  C>T) enquanto o  $\text{SNP}_{RD29-1}$  é uma transversão (A>C).

O polimorfismo  $\text{SNP}_{RD29-6}$  é uma modificação silenciosa já que não modifica o aminoácido aspartato ( $D_{500}$ ) enquanto os outros modificam os aminoácidos codificados (Tabelas 10A e 10B). Assim, os  $\text{SNP}_{RD29-1}$  até  $\text{SNP}_{RD29-3}$  alteram a segunda base do códon, trocando respectivamente o glutamato ( $E_{450}$ ) por alanina, a arginina ( $R_{472}$ ) por histidina e o aspartato ( $D_{483}$ ) por glicina. O  $\text{SNP}_{RD29-3}$ , o  $\text{SNP}_{RD29-4}$  afeta a primeira base do códon, trocando o glutamato

(E<sub>484</sub>) por lisina, enquanto o SNP<sub>RD29-5</sub> troca a terceira base do códon, a metionina (M<sub>498</sub>) é trocada por isoleucina.

As análises das sequências mostram também a presença de modificações pontuais de nucleotídeos que modificam os aminoácidos 471 e 474, respectivamente para os genótipos “Catuaí 25” e “Canephora” (Tabela 10A). No primeiro caso, observa-se uma glicina (G) enquanto todos os outros genótipos apresentam uma arginina e para o segundo caso observa-se, um triptofano enquanto todos os outros genótipos apresentam uma glicina. Como esses resultados foram deduzidos de seqüências realizadas usando o pool do produto de PCR sem clonagem, só será possível conferir essas modificações ao realizar os seqüenciamentos dos produtos de PCR clonados (trabalho atualmente em andamento).

Tabela 9A SNPs do gene *RD29* localizados na seqüência codante. A seqüência de referência do gene *RD29* (contig 5590) é apresentada na tabela acima antes dos genótipos. Para cada um dos genótipos testados, os nucleotídeos divergentes (SNPs) da seqüência de referência são apresentados em vermelho. Nucleotídeos iguais ao da seqüência de referência são apresentados em amarelo. As posições dos SNPs na seqüência de referência são indicadas na última linha da tabela. Os genótipos não analisados por problemas de seqüenciamentos são indicados (R = refazer)

	SNP <sub>RD29-1</sub> (A/C)	SNP <sub>RD29-2</sub> (G/A)	SNP <sub>RD29-3</sub> (A/G)
<b>Gene <i>RD29</i></b>	AG <b>A</b> GG	G <b>C</b> GTC	GG <b>A</b> TG
<b>Genótipos</b>			
01-M. Novo	AG <b>A</b> GG AG <b>C</b> GG	G <b>C</b> GTC GC <b>A</b> TG	GG <b>C</b> TG
02-Catuaí 25	AG <b>C</b> GG	G <b>C</b> GTC	GG <b>C</b> TG
03-Typica	AG <b>C</b> GG	G <b>C</b> GTC	GG <b>C</b> TG
04-Bourbon	AG <b>C</b> GG	G <b>C</b> GTC	GG <b>C</b> TG
05-Rubi	R	R	R
06-159	AG <b>C</b> GG	G <b>C</b> GTC	GG <b>C</b> TG

“Tabela 9A, continua”

"Tabela 9A, conclusão"			
07-E516	R	R	R
08-E464	AGAGG AGCGG	GCGTC	GGGTG
09-E007	R	R	R
10-G237	R	R	R
11-E017	R	R	R
12-E238	AGCGG	GCAATC	GGGTG
13-E123B	AGAGG AGCGG	GCGTC GCAATC	GGGTG
14-E123A	AGAGG	GCAATC	GGGTG
15-UW099	AGAGG	GCAATC GCGTC	GGGTG
16-UW002	R	R	R
18-C3001	AGAGG	GCGTC	GGGTG
19-Canep.	AGAGG	GCGTC	GGATG GGGTG
20-G2011	AGAGG	GCGTC	GGGTG
21-CI007	AGAGG AGCGG	GCGTC GCAATC	GGGTG
22-C4001	R	R	R
23-C.14	R	R	R
24-C.22	R	R	R
25-G2020	R	R	R
26-Psil.	AGAGG	GCGTC GCAATC	GGGTG
27-Racem.	AGAGG	GCAATC	GGGTG
28-Eug.	R	R	R
<b>Posição SNP</b>	<b>1459</b>	<b>1525</b>	<b>1558</b>

Tabela 9B SNPs do gene *RD29* localizados na sequência codante. A sequência de referência do gene *RD29* (contig 5590) é apresentada na tabela acima antes dos genótipos. Para cada um dos genótipos testados, os nucleotídeos divergentes (SNPs) da sequência de referência são apresentados em vermelho. Nucleotídeos iguais aos da sequência de referência são apresentados em amarelo. As posições dos SNPs na sequência de referência são indicadas na última linha da tabela. Os genótipos não analisados por problemas de seqüenciamentos são indicados (R = refazer)

	SNP <sub>RD29-4</sub> (G/A)	SNP <sub>RD29-5</sub> (G/A)	SNP <sub>RD29-6</sub> (T/C)
<b>Gene <i>RD29</i></b>	AT <b>G</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
<b>Genótipos</b>			
01-M. Novo	AT <b>G</b> AA GT <b>A</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG GAT <b>C</b> AG
02-Catuai 25	GT <b>A</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
03-Typica	GT <b>A</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
04-Bourbon	GT <b>A</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
05-Rubi	R	R	R
06-159	GT <b>A</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
07-E516	R	R	R
08-E464	AT <b>G</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG GAT <b>C</b> AG
09-E007	R	R	R
10-G237	R	R	R
11-E017	R	R	R
12-E238	AT <b>G</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG GAT <b>C</b> AG
13-E123B	AT <b>G</b> AA GT <b>A</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG GAT <b>C</b> AG
14-E123A	AT <b>G</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
15-UW099	AT <b>G</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
16-UW002	R	R	R
18-C3001	AT <b>G</b> AA	AT <b>G</b> GT	GAT <b>T</b> AG
19-Canep.	AT <b>G</b> AA	AT <b>G</b> GT AT <b>A</b> GT	GAT <b>T</b> AG

“Tabela 9B, continua”

“Tabela 9B, conclusão”			
20-G2011	ATGAA	ATGGT	GATAG
21-C1007	ATGAA GTAAA	ATGGT	GATAG GACAG
22-C4001	R	R	R
23-C.14	R	R	R
24-C.22	R	R	R
25-G2020	R	R	R
26-Psil.	ATGAA	ATGGT ATAGT	GATAG
27-Racem.	ATGAA	ATGGT	GATAG
28-Eug.	R	R	R
<b>Posição SNP</b>	<b>1560</b>	<b>1604</b>	<b>1610</b>

Tabela 10A Sequências de aminoácido deduzidas das regiões nucleotídicas contendo os SNP localizados na parte codante do gene RD29 (Contig 5590). Para cada um dos genótipos testados, os aminoácidos divergentes da proteína RD29 de referência são apresentados em vermelho. Aminoácidos iguais ao da sequência de referência são apresentados em amarelo. As posições dos aminoácidos referentes aos SNPs na sequência da proteína de referência são indicadas na última linha da tabela. Para o SNPRD29-2, os aminoácidos G471 do Catuaí 25 e W474 do Canephora G21 divergem das outras sequências estão em vermelho. Os genótipos não analisados por problemas de sequenciamentos são indicados (R = refazer)

	SNP <sub>RD29-1</sub>	SNP <sub>RD29-2</sub>	SNP <sub>RD29-3</sub>
<b>Proteína RD29</b>	<b>KEE<b>V</b>VG</b>	<b>ARR<b>L</b>LG</b>	<b>RE<b>D</b>ED</b>
<b>Genótipos</b>			
01-M. Novo	KEE <b>V</b> VG KE <b>A</b> VG	ARR <b>L</b> LG ARR <b>L</b> LG	RE <b>G</b> KD
02-Catuaí 25	KE <b>A</b> VG	AR <b>G</b> RLG	RE <b>G</b> KD
03-Typica	KE <b>A</b> VG	ARR <b>L</b> LG	RE <b>G</b> KD
04-Bourbon	KE <b>A</b> VG	ARR <b>L</b> LG	RE <b>G</b> KD
05-Rubi	R	R	R
06-I59	KE <b>A</b> VG	ARR <b>L</b> LG	RE <b>G</b> KD
07-E516	R	R	R
“Tabela 10A, continua”			

“Tabela 10A, conclusão”

08-E464	KEEYVG KEAYVG	ARRLG	REGED
09-E007	R	R	R
10-G237	R	R	R
11-E017	R	R	R
12-E238	KEAYVG	ARRLG	REGED
13-E123B	KEEYVG KEAYVG	ARRLG ARRLG	REKGD
14-E123A	KEEYVG	ARRLG	REGED
15-UW099	KEEYVG	ARRLG ARRLG	REGED
16-UW002	R	R	R
18-C3001	KEEYVG	ARRLG	REGED
19-Canep.	KEEYVG	ARRLW	REDED REGED
20-G2011	KEEYVG	ARRLG	REGED
21-CI007	KEEYVG KEAYVG	ARRLG ARRLG	REGED
22-C4001	R	R	R
23-C.14	R	R	R
24-C.22	R	R	R
25-G2020	R	R	R
26-Psil.	KEEYVG	ARRLG ARRLG	REGED
27-Racem.	KEEYVG	ARRLG	REGED
28-Eug.	R	R	R
<b>Posição aa</b>	<b>450</b>	<b>472</b>	<b>483</b>

Tabela 10B Sequências de aminoácido deduzidas das regiões nucleotídicas contendo os SNPs localizados na parte codante do gene *RD29* (Contig 5590). Para cada um dos genótipos testados, os aminoácidos divergentes da proteína RD29 de referência são apresentados em vermelho. Nucleotídeos iguais ao da sequência de referência são apresentados em amarelo. As posições dos aminoácidos referentes aos SNPs na sequência da proteína de referência são indicadas na última linha da tabela. Os genótipos não analisados por problemas de seqüenciamentos são indicados (R = refazer)

	SNP <sub>RD29-4</sub>	SNP <sub>RD29-5</sub>	SNP <sub>RD29-6</sub>
<b>Proteína RD29</b>	<b>EEEDA</b>	<b>KGMVD</b>	<b>MVDRL</b>
<b>Genótipos</b>			
01-M. Novo	EEEDA EGKDA	KGMVD	MVDRL
02-Catuaí 25	EGKDA	KGMVD	MVDRL
03-Typica	EGKDA	KGMVD	MVDRL
04-Bourbon	EGKDA	KGMVD	MVDRL
05-Rubi	R	R	R
06-I59	EGKDA	KGMVD	MVDRL
07-E516	R	R	R
08-E464	EEEDA	KGMVD	MVDRL
09-E007	R	R	R
10-G237	R	R	R
11-E017	R	R	R
12-E238	EEEDA	KGMVD	MVDRL
13-E123B	EEEDA EGKDA	KGMVD	MVDRL
14-E123A	EEEDA	KGMVD	MVDRL
15-UW099	EEEDA	KGMVD	MVDRL
16-UW002	R	R	R
18-C3001	EEEDA	KGMVD	MVDRL
19-Canep.	EEEDA	KGMVD	MVDRL
20-G2011	EEEDA	KGMVD	MVDRL
21-CI007	EEEDA EGKDA	KGIVD	MVDRL
22-C4001	R	R	R
23-C.14	R	R	R
24-C.22	R	R	R
25-G2020	R	R	R
26-Psil.	EEEDA	KGIVD	MVDRL
27-Racem.	EEEDA	KGMVD	MVDRL
28-Eug.	R	R	R
<b>Posição aa</b>	<b>484</b>	<b>498</b>	<b>500</b>

### 5.3.3 Discussão das análises dos polimorfismos (SNPs)

Os SNPs em região codante do gene podem afetar a composição da proteína caso ocorra a troca, a inserção ou a deleção de nucleotídeos alterando consequentemente a estrutura primária (sequência), a conformação (estrutura secundária e terciária) e finalmente a função da proteína. Nesse trabalho, as buscas de SNPs para os genes *DREB*, *NAC-RD26* e *RD29* foram feitas usando-se as seqüências delimitadas pelas extremidades 5' e 3' dos contigs (cDNAs) correspondentes (ver Apêndice 1, 2 e 3), que compreende as seqüências codantes de cada um desses genes e também uma parte das seqüências flanqueantes 5' - e 3' transcritas não traduzidas (“UTR”).

No total de 826 pb (comprimento da seqüência entre os primers usados), uma quantidade limitada de SNPs foi encontrada para o gene *DREB* de café. Resultados similares foram também obtidos recentemente para os genes *DREB1* e *DREB2* de vários cereais (NAYAK et al., 2009; WEI et al., 2009), demonstrando a grande conservação desse gene essencial na resposta das plantas ao estresse hídrico. Dos quatro SNPs encontrados na região codante do gene *DREB* (Contig 14421), dois SNPs (SNP<sub>DREB2</sub> e SNP<sub>DREB4</sub>) não modificam o aminoácido codificado. Esses SNPs são considerados mutações silenciosas sem nenhum efeito discernível na estrutura e na função da proteína e consequentemente no fenótipo. Portanto, não impedem que essas modificações atuem na eficiência da tradução ou na estabilidade do RNAm. Os dois outros SNPs (SNP<sub>DREB1</sub> e SNP<sub>DREB3</sub>) modificam a proteína DREB sem afetar o domínio AP2/ERF (aminoácidos V<sub>66</sub> até F<sub>122</sub> da proteína deduzida do Contig 14421), nem tampouco o domínio de fixação para as seqüências DRE (LIU, 2006). No primeiro caso ocorre uma troca na parte N-terminal da proteína, onde a isoleucina (I<sub>6</sub>) é trocada por treonina (T<sub>6</sub>). A isoleucina é um aminoácido hidrofóbico, importante na estabilização da estrutura das proteínas pela

promoção de interações hidrofóbicas, enquanto a treonina é um aminoácido hidrofílico capaz de realizar pontes de hidrogênio. Como mostrado pelos perfis de análise de Kyte-Doolittle ([www.proteinlounge.com](http://www.proteinlounge.com)), essa troca pode modificar as características da sequência N-terminal da proteína. A sequência com I<sub>6</sub> é mais hidrofóbica que aquela com T<sub>6</sub>. Essa sequência N-terminal da proteína DREB é conhecida por participar no “*targeting*” da proteína no núcleo, é possível que a modificação trazida pelo SNP<sub>DREB1</sub> tenha efeito na translocação da proteína. No SNP<sub>DREB3</sub> ocorre a troca de alanina (A<sub>213</sub>) por glicina, que são dois aminoácidos neutros e hidrofóbicos, desempenhando assim a mesma função na estabilização da proteína.

Como mencionado anteriormente, a busca de SNPs realizada com o sequenciamento dos pools de amplicons dos GC permitiu somente a identificação dos polimorfismos, mais não permitiu saber (i) como os haplótipos são sorteados para cada uma das formas alélicas, ou (ii) o número de formas alélicas que existe para cada genótipo analisado. Para responder a essas questões é necessário clonar os produtos de PCR e sequenciar aproximadamente 10 clones recombinantes, para assim aumentar as chances de ter o acesso a todas as formas alélicas existentes. Esse trabalho foi iniciado para os genes *DREB*, *NAC-RD26*, *RD29* e *GCI0* e ainda está em fase de conclusão no laboratório.

Já foram obtidos resultados parciais de sequenciamento do gene *DREB*. Usando-se os resultados dessa análise preliminar, foi possível identificar a existência de 8 haplótipos (Tabela 11). (anotados de DREB-A até - H) para o gene *DREB* que se destaca pelo arranjo das bases dos 7 SNPs (Tabela 5A e 5B). Em todos os casos, os resultados do sequenciamento dos clones isolados confirmam os resultados obtidos com o sequenciamento dos pools dos amplicons.

Tabela 11 Haplótipos encontrados para o gene *DREB*. Para cada SNP, os nucleotídeos divergentes do gene *DREB* de referência são apresentados em vermelho. Quando os nucleotídeos são iguais aos da sequência de referência estes são apresentados em amarelo. nd: sequências não determinadas pelo sequenciamento dos clones isolados

Gene <i>DREB</i>	SNP <sub>DREB1</sub> (T/C)	SNP <sub>DREB2</sub> (C/T)	SNP <sub>DREB3</sub> (C/G)	SNP <sub>DREB4</sub> (G/A)	SNP <sub>DREB5</sub> (T/C)	SNP <sub>DREB6</sub> (-/C)	SNP <sub>DREB7</sub> (T/A)
<i>DREB</i>	CAAT <b>T</b> ATG	AGAC <b>C</b> TCT	GTG <b>C</b> AGT	TTC <b>G</b> AATT	CTGTAGT	TAG <b>-</b> TAG	TAG <b>T</b> ACT
DREB-A	<b>C</b>	<b>C</b>	<b>G</b>	<b>A</b>	<b>T</b>	nd	<b>T</b>
DREB-B	<b>T</b>	<b>C</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>A</b>
DREB-C	<b>C</b>	<b>C</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>A</b>
DREB-D	<b>T</b>	<b>T</b>	<b>G</b>	<b>G</b>	<b>T</b>	nd	<b>T</b>
DREB-E	<b>T</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>C</b>	<b>C</b>	<b>A</b>
DREB-F	<b>C</b>	<b>T</b>	<b>G</b>	G	T	nd	T
DREB-G	<b>C</b>	<b>C</b>	<b>G</b>	G	T	nd	T
DREB-H	nd	<b>C</b>	<b>C</b>	<b>A</b>	<b>T</b>	nd	<b>T</b>

Assim, foi possível confirmar que os clones C3001 e 14 de *C. canephora* são homocigotos para o haplótipo DREB-A, sendo todos os demais genótipos de *C. canephora* heterocigotos para os diferentes haplótipos (Tabela 12). Todos os genótipos de *C. arabica* são heterocigotos para os diferentes haplótipos encontrados. Sendo que os haplótipos DREB-F, G e H somente foram encontrados nas formas alélicas oriundas de *C. arabica* (Tabela 12).

Tabela 12 Haplótipos (DREB-A até -H) encontrados para o gene *DREB* para genótipos de alguns acessos de *C. canephora* e *C. arabica*. As cores usadas para identificar os haplótipos correspondem às sequências definidas na tabela 11

	Genótipos	DREB	DREB	DREB	DREB	DREB	DREB	DREB	DREB
		A	B	C	D	E	F	G	H
<i>C. arabica</i>	M. Novo			Orange	Blue				
	Typica			Orange	Blue			Green	
	Bourbon			Orange	Blue	Pink			
	Rubi		Green	Orange	Blue	Pink			
	Iapar 59	Yellow		Orange	Blue	Pink		Green	
	E516				Blue	Pink		Green	
	E464			Orange	Blue				
	E237			Orange	Blue				Light Blue
	E017			Orange	Blue	Pink	Yellow	Green	
	E238			Orange	Blue				
	E123B				Blue	Pink			
<i>C. canephora</i>	UW099			Orange					
	UW002			Orange					
	C3001	Yellow							
	G2011	Yellow		Orange					
	C4001	Yellow	Green						
	Cl.14	Yellow							
	Cl.22	Yellow	Green	Orange					
<i>Psilantus</i>	Psil.			Orange	Blue	Pink			
<i>C. eugenioides</i>	Eug.				Blue				

Para os genótipos Rubi e I59 de *C. arabica*, foram encontradas as combinações de haplótipos DREB-B/C/D/E e DREB-A/C/D/E/G respectivamente. Isso mostra que nessas plantas tetraploides, o gene *DREB* se encontra na forma heterozigota nos cromossomos oriundos de *C. canephora* e de *C. eugenioides*. É interessante notar que as formas alélicas DREB-A, B e C se encontram também nos acessos de *C. canephora* enquanto as formas alélicas DREB-D e E somente se encontram em *C. arabica*. Desta forma, esses

resultados sugerem que essas duas últimas formas são vinculadas aos cromossomos de *C. eugenioides*. Esses resultados devem ser confirmados pelo seqüenciamento dos clones isolados com o gene *DREB* dessa espécie. Pode-se observar que o clone 14 de *C. canephora* tolerante a seca foi homocigoto do haplótipo A, enquanto o clone 22 suscetível a seca foi heterocigoto, com a presença de três haplótipos (A, B e C). O haplótipo A foi também encontrado no cultivar I59 de *C. canephora*, na forma homocigota. Vale ressaltar que as análises estão em andamento com os outros genótipos, para se identificar e/ou confirmar as formas alélicas existentes.

Para o mesmo gene *DREB*, foram encontrados também três SNPs ( $SNP_{DREB5}$ ,  $SNP_{DREB6}$  e  $SNP_{DREB7}$ ) na parte 3'UTR. Usando-se o programa de análise de estruturas secundárias de RNA, não foi possível notar efeitos drásticos dessas modificações sobre os dobramentos dessa região, que poderiam estar relacionados ao controle da estabilidade do RNAm desse gene.

Para o gene *NAC-RD26*, dos sete SNPs encontrados na região codante do contig 12424, quatro SNPs ( $SNP_{NAC2}$ ,  $SNP_{NAC3}$ ,  $SNP_{NAC6}$  e  $SNP_{NAC7}$ ) não modificam o aminoácido codificado enquanto três SNPs ( $SNP_{NAC1}$ ,  $SNP_{NAC4}$  e  $SNP_{NAC5}$ ) alteram os aminoácidos correspondentes. É possível observar que o  $SNP_{NAC1}$  (G>A) está localizado na seqüência conservada NAM ("No Apical Meristem"; L<sub>14</sub> até L<sub>139</sub>) da proteína codificada pelo Contig 12424 conhecida como domínio essencial e possui função de fator de transcrição (SOUER, 1996). Nenhum desses SNPs estão localizados nas duas seqüências NLS ("nuclear localization signal") dessa proteína (D<sub>77</sub>-R<sub>88</sub> e I<sub>113</sub>-K<sub>129</sub>), nem nos domínios de dimerização que caracterizam as proteínas NAC (OLSEN et al., 2005). Também, é possível notar que os  $SNP_{NAC4}$  e  $SNP_{NAC5}$  são localizados na parte C-terminal da proteína que funciona como um domínio de ativação transcricional (FUJITA et al., 2004). No  $SNP_{NAC4}$  ocorre a troca da S<sub>309</sub> (aminoácido hidrofílico neutro) pela arginina (hidrofílico básico). Essa troca poderá acarretar mudanças na

conformação e/ou na estrutura da proteína RD26, visto que a arginina possui a cadeia lateral maior com um grupo guanidino (-NH<sub>2</sub>), carregado positivamente. Entretanto, quando a S<sub>323</sub> é trocada pelo aminoácido asparagina, como observado no SNP<sub>NAC5</sub>, a mudança conformacional pode não ser significativa, visto que os dois aminoácidos são hidrofílicos e neutros. Como nenhum estudo de mutagênese dirigida e de busca de SNPs no gene *NAC-RD26* foi publicada, não é possível saber se os SNPs identificados nesse trabalho exercem um papel na função da proteína RD-26.

Na sequência total de 1661 pb do gene *RD29*, foram encontrados 6 SNP, 5 deles modificam os aminoácidos correspondentes. A estrutura secundária dessa proteína não parece ser alterada profundamente com essas modificações. Isso pode ser explicado porque nos SNP<sub>RD29-2</sub>, SNP<sub>RD29-4</sub> e SNP<sub>RD29-5</sub>, não ocorrem modificações de caráter hidrofóbico ou hidrofílico dos resíduos trocados. Também é possível notar que nenhum dos SNPs identificados afeta o domínio CAP160 presente na proteína RD29 (I<sub>313</sub>-G<sub>339</sub>) conhecido por estar implicado na estabilização das membranas dos elementos do cito esqueleto e dos ribossomos (pfam07918). Existem muitos trabalhos publicados sobre a expressão do gene *RD29* em plantas e também sobre o uso do seu promotor em transgênicos (NARUSAKA et al., 2003; YAMAGUCHI-SHINOZAKI; SHINOZAKI, 1993). Além disto, muito pouco se sabe sobre a estrutura e a função precisa da sua proteína. A análise da proteína RD29 de cafeeiro (Blastp: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) mostra a existência de vários domínios encontrados também em proteínas de plantas induzidas com o estresse abiótico, principalmente com o frio. Como esses domínios cobrem essa proteína, as modificações dos aminoácidos oriundas dos SNPs identificados (SNP<sub>RD29-1</sub> até SNP<sub>RD29-5</sub>) poderiam afetar a estrutura secundária e terciária da proteína RD29.

## 6 CONCLUSÕES E PERSPECTIVAS

Os resultados de qPCR mostram que as respostas e os perfis de expressão dos GC nas folhas dos cultivares de *C. arabica* podem ser comparáveis em alguns casos com aquelas obtidas previamente em folhas de *C. canephora*. Mostram também a dificuldade em comparar os resultados obtidos com plantas cultivadas em condições controladas de estresse em casa de vegetação com os resultados obtidos com plantas cultivadas no campo. Assim, os nossos resultados mostram que os níveis de estresse hídrico afetam as respostas de expressão dos GC, com aumento da expressão de GC, por exemplo: Os genes *NAC-RD26*, *RD29* e *M6FR*, possuem uma expressão mais intensa em condições severas de estresse hídrico, enquanto em outros genes, *GC10* e *MYB6*, é possível observar uma expressão mais intensa em condições de estresse moderado. Esses resultados podem ser usados em estudos futuros visando-se analisar as respostas moleculares de vários cultivares e variedades em condições de estresse hídrico no campo (projeto em andamento no campo experimental da Embrapa Cerrados, 2010-2012).

Esses resultados mostram também que vários genes apresentaram respostas de expressão diferencial com as condições de estresse, e entre os cultivares I59 e Rubi, como é o caso, por exemplo, do gene *RD29*. Para os genes que codificam fatores de transcrição ou de função não conhecida, fica difícil confirmar essas variações de expressão com testes bioquímicos e fisiológicos. Entretanto, é possível realizar esses testes para os genes que codificam para enzimas ou proteínas implicadas na fotossíntese tais como os genes *M6FR*, *AC*, *OEE* e *RBCS*, por exemplo.

Praxedes et al. (2006) não mostraram uma participação dos açúcares no processo da tolerância a seca nas folhas de *C. canephora* var. Conilon. Entretanto, os resultados obtidos para o gene *M6FR*, mostram uma alta

expressão desse gene com o estresse hídrico no cultivar I59, sugerindo que seria interessante avaliar os teores do manitol nas folhas das plantas I e NI dos cultivares I59 e Rubi analisados nesse estudo. Análises dos teores de ácidos clorogênicos podem ser correlacionados com as variações de expressão do gene *CCoAMT* que foram observadas.

Os resultados de expressão dos genes *AC* e *OEC* mostram uma expressão elevada (no ano 2009) com o estresse hídrico no cultivar I59, sendo pertinente se avaliar os parâmetros fotossintéticos (*A*, taxa de assimilação de CO<sub>2</sub>; *g<sub>s</sub>*, condutância estomática; *E*, taxa de transpiração; *A/E*, eficiência do uso da água [“WUE: water use efficiency”]) desses cultivares e compará-los com as variações de expressão dos GC observadas. Adicionalmente a testes ainda em andamento no laboratório para se estudar a expressão dos genes *RBCS* codificando as isoformas da RUBISCO, esses parâmetros fotossintéticos poderiam auxiliar na compreensão dos perfis de expressão obtidos durante esse estudo. Todas essas medidas fisiológicas foram também realizadas no âmbito desse projeto e estão agora em fase de análise dos dados.

Contudo, os resultados obtidos pela identificação de SNPs nesse trabalho correspondem ao primeiro passo para a identificação de marcadores funcionais que podem ser desenhados especificamente para todas as formas alélicas desses GC (*DREB*, *NAC-RD26* e *RD29*). Mas ainda há a necessidade de se estabelecer corretamente os haplótipos (formas alélicas) presentes para cada gene, e se determinar para cada gene os genótipos das plantas de café analisadas.

Para que isso ocorra, é necessário finalizar a clonagem e o seqüenciamento dos clones isolados para cada um dos GC, trabalho em andamento em nosso laboratório.

Os dados gerados até o momento juntamente com os resultados que surgirão após o seqüenciamento dos clones abrirá a possibilidade de saber quais são as combinações alélicas existentes para as espécies, cultivares e clones de

*Coffea* que foram analisadas e também para verificar se existem formas alélicas que estão associadas com as regiões geográficas de origem que foram selecionadas (positivamente ou negativamente) ao longo dos programas de melhoramento. Uma vez realizado esse tipo de análise, será possível saber se algumas formas alélicas podem estar associadas como o fenótipo (quando este é conhecido) das plantas com relação à tolerância a seca.

Os resultados apresentados para a busca e análise de SNPs para os genes *DREB*, *NAC-RD26* e *RD29* mostram que existe uma quantidade limitada de modificações nas regiões codantes desses genes. Isso pode ser explicado pela existência de uma pressão de seleção muito forte sobre esses genes ao longo da evolução relacionada com a importância das proteínas correspondentes na cascata de eventos na qual elas estão envolvidas. Além da importância desses genes nos mecanismos de resposta das plantas ao estresse abiótico, decidimos buscar e analisar os SNPs com base nos resultados de expressão, (obtidos no laboratório), mostrando a expressão diferencial dos genes com o estresse hídrico em folhas de diferentes clones de *C. canephora* Conilon. Assim, os desenvolvimentos dos projetos de sequenciamento de DNA genômico de café irão facilitar o acesso a esse tipo de sequências e também as sequências promotoras que controlam a expressão dos genes em nível transcricional (eficiência da iniciação da transcrição).

Na ausência do acesso a essas sequências, a discussão sobre a parte de busca e análise dos SNPs foi baseada essencialmente sobre as possíveis consequências dessas modificações no nível das proteínas. Uma vez realizados os sequenciamento das formas alélicas e conhecidos os haplótipos que existem para cada GC, será mais fácil tentar prever os efeitos dessas modificações no nível bioquímico, por exemplo, usando os programas de modelagem 3D (RASMOL) de proteínas para verificar se as modificações nos aminoácidos alteram ou não alteram a conformação das proteínas.

## REFERÊNCIAS

ALVES, G. S. C. et al. Análise quantitativa da expressão de genes candidatos em genótipos de *Coffea arabica* submetidos ao déficit hídrico. In: CONGRESSO DE PÓS-GRADUAÇÃO DA UFLA, 18., 2009, Lavras. **Anais...** Lavras: UFLA, 2009. 1 CD-ROM.

ANDERSEN, J. R.; LÜBBERSTEDT, T. Functional markers in plants. **Plant Science**, Denmark, v. 8, n. 11, p. 554-560, Nov. 2003.

ANDRADE, A. C. Projeto genoma café: uma ferramenta para acelerar o melhoramento de variedades. **Irrigação e Tecnologia Moderna**, Belo Horizonte, v. 73, p. 19-20, mar. 2007.

ANDRADE, A. C. et al. Identification and functional characterization of genes involved in drought stress responses in coffee plants. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 21., 2006, Montpellier. **Proceedings...** Montpellier: ICCS, 2006. p. 125.

ANGELOCCI, L. R. **Água na planta e trocas gasosas/energéticas com a atmosfera:** introdução ao tratamento biofísico. Piracicaba: FEALQ, 2002. p. 71-104.

ANTUNES, R. C. B. et al. Influência da fertirrigação com nitrogênio e potássio nos componentes vegetativos do cafeeiro Arábica em formação. In: SIMPÓSIO BRASILEIRO DE PESQUISA DOS CAFÉS DO BRASIL, 2., 2000, Poços de Caldas. **Anais...** Poços de Caldas: SBC, 2000. p. 802-806.

ASSAD, E. D. **Aquecimento global pode mudar o mapa de produção.** Disponível em: <<http://www22.sede.embrapa.br/cafe/noticias/n30102009.html>>. Acesso em: 17 out. 2009.

ASSAD, E. D. et al. Zoneamento agroclimático para a cultura do café no Estado de Goiás e Sudoeste do Estado da Bahia. **Revista Brasileira de Agrometeorologia**, Passo Fundo, v. 9, n. 3, p. 510-518, ago. 2001.

BARKER, G. et al. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using auto SNP. **Plant Physiology**, Netherlands, v. 132, n. 5, p. 84-91, May 2003.

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

BERTRAND, B. et al. Genetic study of *Coffea canephora* coffee tree resistance to *Meloidogyne incognita* nematodes in Guatemala and *Meloidogyne* sp. nematodes in El Salvador for selection of rootstock varieties in Central America. **Euphytica**, Netherlands, v. 113, n. 2, p. 79-86, Apr. 2000.

BORÉM, A.; CAIXETA, A. **Marcadores moleculares**. 2. ed. Viçosa, MG: Folha de Viçosa, 2009. 374 p.

BOUHARMONT, P.; AWEMO, J. La sélection végétative du caféier au Cameroun. **Café Cacao Thé**, Paris, v. 23, n. 4, p. 227-254, 1979.

BRAY, E. Molecular responses to water deficit. **Plant Physiology**, Washington, v. 103, n. 4, p. 1035-1040, Dec. 1993.

CAIXETA, E. T. et al. Microsatellite markers from the Brazilian Coffee Genome Project. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 22., 2008, Campinas. **Anais...** Campinas: UNICAMP, 2008. 1 CD-ROM.

CARVALHO, C. H. S. **Cultivares de café: origem, características e recomendações**. Brasília: EMBRAPA Café, 2008. 334 p.

CATTANEO, L. F. **Avaliação da divergência genética e análise de gerações em mamoeiro (*Carica papaya* L.)**. 2001. 95 p. Tese (Doutorado em Produção Vegetal) - Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, 2001.

CHEN, T. H. H.; MURATA, N. Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. **Plant Biology**, Oregon, v. 5, n. 2, p. 250-257, June 2002.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Avaliação da safra agrícola cafeeira 2009, segunda estimativa maio/2009**. Disponível em: <<http://www.conab.gov.br>>. Acesso em: 18 dez. 2009.

DAMATTA, F. M. et al. Drought tolerance of two field-grown clones of *Coffea canephora*. **Plant Science**, Shannon, v. 164, n. 1, p. 111-117, Jan. 2003.

DAMATTA, F. M.; RAMALHO, J. D. C. Impacts of drought and temperature stress on coffee physiology and production: a review. **Brazilian Journal of Plant Physiology**, Londrina, v. 18, n. 1, p. 55-81, Jan./Mar. 2006.

DUBOUZET, J. G. et al. Os DREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought, high-salt and cold responsive gene expression. **The Plant Journal**, Tokyo, v. 33, n. 4, p. 751-763, Feb. 2003.

DUSSERT, S. et al. Coffee (*Coffea canephora*). In: \_\_\_\_\_. **Genetic diversity of cultivated tropical plants**. Montpellier: CIRAD, 1999. p. 239-258.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Projeto genoma café**. Disponível em: <<http://www.cenargen.embrapa.br/biotec/genomacafe/projeto.html2002>>. Acesso em: 23 out. 2009.

FAHL, J. I. et al. Exchange, growth, yield and beverage quality of *Coffea arabica* cultivars grafted on to *C. canephora* and *C. congensis*. **Experimental Agriculture**, Cambridge, v. 37, n. 2, p. 241-252, May 2001.

FAZUOLI, L. C. Cultivares de café arábica do IAC: um patrimônio da cafeicultura brasileira. **O Agrônomo**, Campinas, v. 59, n. 1, p. 12-15, mar. 2007.

FAZUOLI, L. C. et al. Avaliação antecipada de progênies e de plantas matrizes no café Robusta. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 20., 2000, Poços de Caldas. **Anais...** Brasília: EMBRAPA Café; Minasplan, 2000. p. 462-464.

FERRÃO, R. G. et al. **Robustão Capixaba, variedade clonal de café conilon tolerante à seca**. Vitória: UFES, 1999. 10 p. (Comunicado Técnico, 98).

FUJITA, M. et al. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. **The Plant Journal**, Denmark, v. 39, n. 10, p. 863-876, Oct. 2004.

GEHRIG, H. H.; ROSICKER, H.; KLUGE, M. Detection of DNA polymorphisms in the genus *Kalanchoe* by RAPD-PCR fingerprint and its relationships to infrageneric taxonomic position and ecophysiological photosynthetic behaviour of the species. **Plant Science**, Berlin, v. 125, n. 1, p. 41-45, June 1997.

GONÇALVES, W. et al. Cultivar Icatu Vermelho Iac-925: fonte de resistência a nematóides do gênero meloidogyne. **Genetics and Molecular Biology**, Ribeirão Preto, v. 21, n. 5, p. 228-230, 1998.

GRIVET, L. et al. ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. **Theoretical and Applied Genetics**, Berlin, v. 106, n. 2, p. 190-197, Jan. 2003.

INSTITUTO AGRONÔMICO DO PARANÁ. **Projeto do Programa Café que tem participação do IAPAR é premiado pela Embrapa**. Curitiba, 2009.

Disponível em:

<<http://www.iapar.br/modules/noticias/article.php?storyid=716>>. Acesso em: 29 out. 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção agrícola mundial culturas temporárias e permanentes 2009**. Disponível em: <[http://www.ibge.gov.br/home/estatistica/economia/pam/2009/PAM2009\\_Publicacao\\_completa.pdf](http://www.ibge.gov.br/home/estatistica/economia/pam/2009/PAM2009_Publicacao_completa.pdf)>. Acesso em: 20 mar. 2010.

INTERNATIONAL COFFEE ORGANIZATION. **As estatísticas do comércio do café**. Disponível em: <<http://www.ico.org>>. Acesso em: 25 nov. 2009.

KANG, J. et al. Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. **Plant Cell**, Oxford, v. 14, n. 1, p. 343-357, Feb. 2002.

KAUL, S. et al. The arabidopsis genome initiative. **Nature Genetics**, New York, v. 408, n. 6814, p. 796-815, Dec. 2000.

KUMAR, L. S. DNA markers in plant improvement: an overview. **Biotechnology Advances**, New Delphi, v. 17, n. 2/3, p. 143-182, Sept. 1999.

LARCHER, W. **Physiological plant ecology: ecophysiology and stress physiology of functional groups**. New York: Springer, 2003. 297 p.

LASHERMES, P. Inheritance and genetic mapping of self: incompatibility in *Coffea canephora* Pierre. **Theoretical and Applied Genetics**, Berlin, v. 93, n. 3, p. 458-462, Jan. 1996.

LASHERMES, P.; ANDRADE, A. C.; ETIENNE, H. Genomics of coffee, one of the world's largest traded commodities. **Genomics of Tropical Crop Plants**, New York, v. 1, n. 1, p. 203-226, May 2008.

LASHERMES, P. et al. Molecular characterization an origin of the *Coffea arabica* L. **Genome**, Ottawa, v. 26, n. 2, p. 259-266, Mar. 1999.

LEROY, T. et al. Genetics of coffee quality. **Brazilian Journal of Plant Physiology**, Londrina, v. 18, n. 1, p. 229-242, Jan./Mar. 2006.

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

LIU, N. Cloning and functional characterization of Ppdbl1 gene encoding a DRE-binding transcription factor from *Physcomitrella patens*. **Planta**, Beijing, v. 226, n. 4, p. 827-838, Sept. 2007.

LIU, Y. The conserved Ala37 in the ERF/AP2 domain is essential for binding with the DRE element and the GCC box. **Behalf of FEBS**, Beijing, v. 580, n. 5, p. 1303-1308, Feb. 2006.

MARRACCINI, P. Study of drought tolerance mechanisms in coffee plants by an integrated analysis. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 22., 2008, Montpellier. **Proceedings...** Montpellier: ASIC, 2008. p. 874-881.

MARRACCINI, P. et al. Identificação de genes com expressão diferencial em folhas de cafeeiro *coffea canephora* e *coffea arabica* submetidas à diferentes condições de estresse hídrico. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 6., 2009, Vitória. **Anais...** Vitória: UFES, 2009. p. 324-328.

\_\_\_\_\_. Rubisco small subunit of *Coffea arabica*: cDNA sequence, gene cloning and promoter analysis in transgenic tobacco plants. **Plant Physiology and Biochemistry**, Paris, v. 41, n. 1, p. 17-25, July 2003.

MATIELLO, J. B.; DANTAS, S. F. de A. Desenvolvimento do cafeeiro e do sistema radicular, com e sem irrigação, em Brejões, PE. In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEEIRAS, 14., 1987, Campinas. **Anais...** Campinas: UNICAMP, 1987. p. 165.

MENDES, A. N. G.; GUIMARÃES, R. J.; SOUZA, C. A. S. Classificação botânica origem e distribuição geográfica do cafeeiro. In: GUIMARÃES, R. J. et al. (Ed.). **Cafeicultura**. Lavras: UFLA/FAEPE, 2002. p. 39-99.

MONTAGNON, C. Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de *Coffea canephora* Pierre. **Cirad**, Montpellier, v. 42, n. 7, p. 133, 2000.

MONTAGNON, C.; LEROY, T.; YAPO, A. Diversité génotypique et phénotypique de quelques groupes de caféiers (*Coffea canephora* Pierre) en collection. **Café Cacao Thé**, Paris, v. 36, n. 3, p. 187-197, 1992.

NARUSAKA, Y. et al. Interaction between two *cis*-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses. **The Plant Journal**, Tokyo, v. 34, n. 2, p. 137-148, Apr. 2003.

NAYAK, S. N. et al. Isolation and sequence analysis of DREB2A homologues in three cereal and two legume species. **Plant Science**, Shannon, v. 177, n. 5, p. 460-467, Nov. 2009.

OLSEN, A. N. et al. Nac transcription factors: structurally distinct functionally diverse. **Plant Science**, Denmark, v. 10, n. 2, p. 79-87, Feb. 2005.

PINHEIRO, H. A. 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.

PRAXEDES, S. C. et al. Effects of long-term soil drought on photosynthesis and carbohydrate metabolism in mature robusta coffee (*Coffea canephora* Pierre var. *kouillou*) leaves. **Environmental and Experimental Botany**, Oxford, v. 56, n. 3, p. 263-273, July 2006.

RAFALSKI, J. A. Novel genetic mapping tools in plants: SNPs and LD-based approaches. **Plant Science**, Shannon, v. 162, n. 3, p. 329-333, Mar. 2002.

RAMOS, H. J. O. et al. Differential protein expression analysis of *Coffea canephora* plants under drought stress. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE BIOQUÍMICA E BIOLOGIA MOLECULAR, 37., 2008, Águas de Lindóia. **Anais...** Águas de Lindóia: SBBq, 2008. 1 CD-ROM.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chainterminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 74, n. 12, p. 5463-5467, June 1977.

SANTOS, A. S. **Marcadores de DNA no melhoramento genético do feijoeiro (*Phaseolus vulgaris* L.) visando resistência à *Xanthomonas axonopodis* pv. *phaseoli***. 2000. 140 p. Tese (Doutorado em Produção Vegetal) - Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, 2000.

SANTOS, R. C. et al. Assesment of genetic diversity on a sample of cocoa accessions resistant to witches' broom disease based on RAPD and pedigree data. **Bragantia**, Campinas, v. 64, n. 3, p. 361-368, jun. 2005.

SERA, T. Coffee genetic breeding at IAPAR. **Crop Breeding and Applied Biotechnology**, Londrina, v. 1, n. 2, p. 179-190, Sept. 2001.

SHAO, H. B. et al. Understanding molecular mechanism of higher plant plasticity under abiotic stress. **Biointerfaces**, Beijing, v. 54, n. 1, p. 37-45, Jan. 2007.

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Gene expression and signal transduction in water-stress response. **The Plant Physiology**, Tokyo, v. 115, n. 4, p. 327-334, Apr. 1997.

\_\_\_\_\_. Gene networks involved in drought stress response and Tolerance. **Journal of Experimental Botany**, Oxford, v. 58, n. 2, p. 221-227, Jan. 2007.  
SILVA, E. A.; MAZZAFERA, P. Influences of temperature and water in the coffee culture. **The Americas Journal of Plant Science and Biotechnology**, New York, v. 2, n. 1, p. 32-41, May 2008.

SILVA, V. A. **Caracterização fisiológica da contribuição relativa do sistema radicular e da parte aérea na tolerância à seca em *Coffea canephora***. 2007. 68 p. Tese (Doutorado em Fisiologia Vegetal) - Universidade Federal de Viçosa, Viçosa, MG, 2007.

SOARES, R. S. et al. Estudo comparativo de fontes de nitrogênio e potássio empregados na fertirrigação do cafeeiro. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 1., 2000, Poços de Caldas. **Resumos...** Poços de Caldas: SBC, 2000. v. 2, p. 852-855.

SOMERS, D. J. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. **Genome**, Ottawa, v. 49, n. 3, p. 430-437, June 2003.

SOUER, E. The no apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. **Cell**, Netherlands, v. 85, n. 2, p. 159-170, Apr. 1996.

TADA, Y.; TAKAOKI, K. Proteomic analysis of salt-responsive proteins in the mangrove plant, *Bruguiera gymnorhiza*. **Plant Cell Physiology**, Oxford, v. 50, n. 3, p. 439-446, Jan. 2009.

TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. 3. ed. Porto Alegre: Artmed, 2004. 715 p.

TRAN, L. S. P. et al. Coexpression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in Arabidopsis. **The Plant Journal**, Tokyo, v. 49, n. 1, p. 46-63, Jan. 2007.

USECHE, F. J. High-Throughput identification, database storage and analysis of SNPs in EST sequences. **Genome Informatics Series**, Newark, v. 12, n. 12, p. 194-203, Dec. 2001.

VARSHNEY, R. K.; GRANER, A.; SORRELLS, M. E. Genic microsatellite markers in plants: features and applications. **Biotechnology**, Berlin, v. 23, n. 1, p. 48-55, Jan. 2004.

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

VINECKY, F. Molecular characterization of candidate genes involved in drought resistance in coffee plants. In: INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 22., 2008, Campinas. **Anais...** Campinas: UNICAMP, 2008. 1 CD-ROM.

WANG, W.; VINOCUR, 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.

WEI, B. et al. Assaying single nucleotide polymorphism in wheat (*Triticum aestivum* L.) with allele-specific PCR. **Science Agriculture Sinica**, Beijing, v. 39, n. 2, p. 1313-1320, Sept. 2006.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Characterization of the expression of a desiccation-responsive *rd29* gene of *Arabidopsis thaliana* and analysis of its promoter in transgenic plants. **Molecular and General Genetics**, Tokyo, v. 236, n. 2/3, p. 331-340, Feb. 1993.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. **Trends in Plant Science**, Denmark, v. 10, n. 1, p. 88-94, Feb. 2005.

YANAGUI, K. et al. Análise in vivo da diversidade nucleotídica de *Coffea* spp. In: CONGRESSO BRASILEIRO DE GENÉTICA, 55., 2009, Águas de Lindóia. **Anais...** Águas de Lindóia: SBG, 2009. 1 CD-ROM.

YU, J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). **Science**, New York, v. 296, n. 5565, p. 79-92, Apr. 2002.

ZHIFANG, G.; LOESCHER, W. H. Expression of a celery mannose 6-phosphate reductase in *Arabidopsis thaliana* enhances salt tolerance and induces biosynthesis of both mannitol and a glucosyl-mannitol dimmer. **Plant, Cell and Environment**, Oxford, v. 26, n. 2, p. 275-283, Feb. 2003.

## APÊNDICE

### APÊNDICE A - Sequência do gene *DREB* com os polimorfismos associados

GATCCCCCGGGCTGCAGCATTAACTGCAATTATATATATATACCCCTCTTTTTGAGTTTTACA  
 ACACCAAGTCGTACGATTCCCTCGCTTCAGCCTTTAACTTTTCAGTTGAATTAACCTCCTCA  
 CTGTCCACTACTACTACTATTTTTTCAAGAATGAACCATTTCTCAAATATGCTACTCCGA  
 CCCAATTCCCTAATCCATCATCCCCTGATTTTTCCAGACTCTTTCATCAACTTCCGACAACG  
 TAGTCAATGCCTGCAAAGCCAATTACTCGGACGAAGAAGTGCTCTTAGCTTCCAATAAT  
 CCCAAGAAAACGTGCAGGCAGGAAAAAGTTTCGCGAAAACAGGCATCCCGTGACAGGGG  
 CATCAGGAGAAGGAACTCTGGCAAGTGGGTCTGTGAAATCAGAGAACCACAAGAAAT  
 CAAGGATATGGCTGGGCACATTCCTACTGCGGAAATGGCTGCTAGAGCCCATGACGTG  
 GCGGCTATTGCCCTCAGAGGCCGTTCTGCTTGTCTGAACTTTGCTGACTCTGCTTGGAG  
 GCTGCCCGTTCCGGAATCCCCGAGCCTAAGCACATTCAGAAGGCTGCTGCCGAGGCCG  
 CTGAGGCCCTTTAGGCCGTCGGAGTCGTGTGATGGGGTATCATCAGGTGCCCTCTGGTGAT  
 CATGCTAATAAAGAGGAATGTGGATCAGTGGCATTCCCGGAAAATGTTTTCTTCATGGA  
 TGAGGAGGCGGTTTTTTGGCATGCCTGGCTTGATTGCCAATATGGCAGAAGGATTGATGC  
 TGGCTCCACCCCAATGTGCAGTTGGGGATGACCTCGAACTTGAAGCTAATGCTGATATG  
 TTTCTATGGAGCTATTCGATTTGATTTACTGGTGATATGTAGAAGTACTCTGTAGTTGG  
 TGGATCGTAGTAG()TAGTACTTTTTGCTAGTACTAGACTACTAGCTTGCTACGAACTG  
 CGTAGTAAATTATCTATTCCGTACTGCAGTTTTTTGGGGACCTGTACGGATAGTGTGGAA  
 CGCAATGTACAAAACAAAAGTAAGAGTAGGTTGTACAGTTAGTTAAATAACTACTGCCA  
 AGAGAGCACTTCTTGTGTTCTATGTAAGTGTGAAAGTAATGAAAGATTACCACCATTT  
 GCCTGATGACAGGTAAGTGTGTAATGAATGTTTCGTGATGTGAAAATAAGGTATTAA  
 AANAAAAAAAAAAAAAAAAAAAA

Legenda: Sequência nucleotídica do gene *DREB* de referência

A sequência (1205pb) referente ao Contig14421 contém os SNPs (SNP<sub>DREB1</sub> até SNP<sub>DREB7</sub>) que são identificados com cores diferentes (SNP<sub>DREB1</sub> **I**, SNP<sub>DREB2</sub> **C**, SNP<sub>DREB3</sub> **C**, SNP<sub>DREB4</sub> **G**, SNP<sub>DREB5</sub> **T**, SNP<sub>DREB6</sub> **O** e SNP<sub>DREB7</sub> **T**). A sequência codante (149-850) é indicada em negrito. Os primers *DREB1A-F*: 5'GTTGAATTAACCTCACTGTCCACTA3' e *DREB1A-R*: 5'CCAAAACTGCAGTACGGAATAGA3' usados para a amplificação do gene estão sublinhados.

### Sequência da proteína DREB deduzida do Contig14421

MNHFS **I** CYSDPIPNPSSP **D** FPDSSSTSDNVVNACKANYSDEEVLLASNNPKKRAGRKKF  
 RETRHP **VYRGIRRRNSGKWVCEVREPNKKSRIWLGTFPTAEMAARAHDVAAIALGRSA**  
**CLNF**ADSAWRLPVPESEPEPKHIQKAAAEAAEAFRPSESCDGVSSGASGDHANKEECGSV  
 AFPENVFFMDEEAVFGMPGLIANMAEGLMLPPPQCA **A** VGDDLELEANADMFLWSY **S** I

Legenda: Sequência da proteína putativa *DREB*

A sequência (233 aminoácidos) da proteína putativa *DREB* é apresentada junto com os aminoácidos relacionados com os SNP identificados: SNP<sub>DREB1</sub> (**I**), SNP<sub>DREB2</sub> (**D**), SNP<sub>DREB3</sub> (**A**) e SNP<sub>DREB4</sub> (**S**). O domínio AP2/ERF (aminoácidos V<sub>66</sub> até F<sub>122</sub>) é indicado em negrito.

## APÊNDICE B - Sequência do gene *NAC-RD26* com os polimorfismos associados

TTCTGGGAGAGCGTTTTCCAGAGTTTCGACTTCTCCAACCGTCCAACGCCACCTCCC GCC  
 CTTTTTTGGCCCTGCGCTCTAGTCAAAGTTATTTTTTAACTATTTTCTTCTGAAAGCA  
 AAAGGGAAAAAAAAAAAAATGGGTGTTTCGAGAACTGACCCGCTTTTCAAACTGAGCTT  
 GCCACCTGGGTTTAGATTCTATCCTACAGACGAGGAGCTGTTGGTGCAGTACCTCTGCA  
 GAAAAGTTGCAGGCCATGATTTTAATCTTCAAATTATAGGAGAGATTGATCTTTACAAG  
 TTTGACCCATGGGATCTTCCGAGTAAGGCGATATTTGGGGAGAAAGAATGGTACTTTTT  
 TAGCCCCAGGGACAGGAAGTACCCGAATGGATCGAGACCCAATAGAGTTGCTGGCTCCG  
 GCTACTGGAAAGCCACTGGAACCGACAAGGTCATCACCACCGAAGGACGTAAAGTTGGG  
 ATCAAGAAAGCCCTCGTTTTTTATGTGGGCAAAGCACCTAAAGGCACCAAGACCAATTG  
 GATCATGCATGAATACAGACTCTCCGAGCCTCAAAGAAAAAATGGTAGCGCCAGGTTGG  
 ATGATTGGGTGCTATGTGCAATTTACAAGAAAAATTCAAGTGCAGGTGCAAAGCCGGTT  
 TCTGGGCTTCAGAGCAGAGAGCACAGTCATGGCTCATCTACGTCATCCTCATCTCAGTT  
 CGACGATGTTCTCGAGTCGTTACCGGAGATTAGTGACCGGTTTTTTCGCTTACCAAGAA  
 TGAATCTCTCAAGAATCTCCACCAAGATGATCAGAAAATCAATATTCAGAACTTGGGT  
 TCAGGGAGCTTTGATTGGGCGACCTCGCCGGACTGAATTCGTTGCCGGAACCTCGGTCC  
 CGGAAGCCAAGCTCAAGCTATTCCGGCAGCAGCTGCACATGGACATGTGAACAGCAACG  
 TGGCCAGCTGCAATAACAACGATAGGAATAATAATGTGAATGCCCAGAATGACATGTAC  
 GTCCCATCGTTTTCCGCAAGCCGGCCACGTGGACGAGGAAGTACAGAGGGGCTGAGGAG  
 TACACCGGGTTGAGAATTCGAGTTTTCTTTCAGCATAACGCCAGCCCCTTGTGCCTCACA  
 CACAGCTTTTCAAACTCGGTGGACCCATTCCGGATCCGGTACCCGCCGACCCAAACC  
 GGGATTATGGGTTTTAGGCCATGAAGAGTAGATTTAGGAGATAAAAATCATGGGTGTA  
 AACTGAAGATTCTGGTTTTGGGGTGAATTTGTACTATATGGATTTCTTTGGGCCGACAA  
 TTTGACTTTTAATTC

Legenda: Sequência nucleotídica do gene *NAC-RD26* de referência

A sequência (1319pb) referente ao Contig12424 contém os SNPs (SNP<sub>NAC1</sub> até SNP<sub>NAC7</sub>) que são identificados com cores diferentes (SNP<sub>NAC1</sub> G, SNP<sub>NAC2</sub> A, SNP<sub>NAC3</sub> C, SNP<sub>NAC4</sub> T, SNP<sub>NAC5</sub> G, SNP<sub>NAC6</sub> C e SNP<sub>NAC7</sub> G). A sequência codante (137-1204) é indicada em negrito. Os *primers NAC-RD26-F*: 5'AAAAAAAATGGGTGTTTCGAGAAACT3' e *NAC-RD26-R*: 5'CTACTCTTCATGGCCTAAAACCCAT3' usados para a amplificação do gene estão sublinhados.

### Seqüência da proteína NAC-RD26 deduzida do Contig12424

MGVRETDP L S Q L S L P P G F R F Y P T D E E L L V Q Y L C R K V A G H D F N L Q I I **G** E I D L Y K F D P W D L  
**P S K A I F G E K E W Y F F S P R D R K Y P N G S R P N R V A G S G Y W K A T G T D K V I T T E G R K V G I K K A L V**  
**F Y V G K A P K G T K T N W I M H E Y R L S E P Q R K N G S A R L D D W V L C R I Y K K N S S A G A K P V S G L Q S R**  
 E H S H G S S T S S S S Q F D D V L E S L P E I S D R F F A L P R M N S L K N L H Q D D Q K I N I Q N L G S G S F D W  
 A T L A G L N S L P E L G P G S Q A Q A I P A A A A H G H V N S N V A S C M N N D R M N N V N A Q N D M Y V P S F P **Q**  
 A G H V D E E V Q S G L R **S** H R V E N S S F F Q H N A **S** P L V P H **T** H S F H N S V D P F G I R Y P P T Q T G I M G F R  
 P

Legenda: Seqüência da proteína putativa *NAC-RD26*

A seqüência (355 aminoácidos) da proteína putativa NAC-RD26 é apresentada junto com os aminoácidos relacionados com os SNP identificados:

SNP<sub>NAC1</sub> **G**, SNP<sub>NAC2</sub> **Q**, SNP<sub>NAC3</sub> **S**, SNP<sub>NAC4</sub> **S**, SNP<sub>NAC5</sub> **S**, SNP<sub>NAC6</sub> **P** e SNP<sub>NAC7</sub> **T**. O domínio NAM (“No Apical Meristem”; L<sub>14</sub> até L<sub>139</sub>) é indicado em graxo enquanto as duas seqüências NLS (“nuclear localization signal”; D<sub>77</sub>-R<sub>88</sub> e I<sub>113</sub>-K<sub>129</sub>) são subscritas.

### APÊNDICE C - Sequência do gene *RD29* com os polimorfismos associados

TGTCAAAAGCAAAATCAAGCATTTGTCTGAAACAAAAACATAGTTAAAGAGAGTTGTTT  
 TGGTTAAGAATTTGTGTGTTTTTTGTAGAAAAGAAAGTTTGTGGAAACAGGCATGGAATC  
 CCAATTACACCGCCCTTATGACCATACTGATGATCAAGATCCCCAACACGCAGCGGTTG  
 AGGATGAAGGAGACCATCATCATCATGAGAAGACATCAGTGATAAGGAAAGTGAAGGCG  
 AAAGCGAAGAAGATCAAGGAGACTCTTGCCAAGCATGGACTTGGTTATGAACATGAACA  
 TGAACGAGATTATAGTCATGATGATCCTGATGAAGGTGAAGAAAATGAGGATGAAGAAA  
 TGGAAGAAGACCCCTGAAGTCCATGGTGCACCCATGTATGAGTCGACTGTGATTGGAAGT  
 GCAATTCCTGCACAAAATGTCAATTTGGAGAAAACCAACAGCAATAGGGGAAGATAGCTA  
 TGAAGGTCAGAGGAATGCTGGGAAAACAGGCGCCACAAGGCCTGCTTTAGAGGGGCAAT  
 CAGGGGAGAATCTGGGAAAACCATCAGCCATGGAAGTGGTTCATGCGCCTGAGGACAAG  
 GATATATCATTTCCCCCTGAAATTCGTGAAACAAAAGGGGATTGAGGTGCCAGAAAACGA  
 CAATGAAAATGTTGGCCCTCAGGGGTTTAGAATTGGTGCTCTAGAAGGCTTAGAGGAAG  
 ATCCTCAAGCGCCGAAAAACCGGCCAGGAGAAGTTCGCCTTCAAATTATCAAATAAAA  
 GTTACTGATCCAACCTGAAAGGGTGGTGAAGAAGTTGGAATAACCCCACTGATCCAATC  
 CTTCAATAAGATGGGTGTTTATGATGAATCAGTGCCTAAATCGGAATCAGAACAAGAGG  
 CATATACAGGAAGCCATGGTCAGTTTGCTCCAGAGCCAAATCCTACTGAAGGCAAATCT  
 GATTGAGCTCCAAAAGTTGTGATCCTAGCAAACCAGAAGACAACCTTACCGCGTGACAC  
 ATTAACCGGAAAATTATCAGACCAGAGCGGTTATGTTGAGAAGATATCACTAGCTACAT  
 CCACTATTGCTGACAAAAGCAATCTCTGCCAAGAATGTAGTCGCTTCCAAGCTTGGATAT  
 GGAGGCACTGAAGGTGGTAAAGTCCCTGAAAACCGATGAAAAAAGAATGCTGCAAAATC  
 AGGTGCATCCCCAGCGGAGTATGCACATAAAGTCTCAGCGACAGTTACTGACAAACTGG  
 CACCAGTTTACCAGAAAAGTCGCAGATGCTGGAAGTTCTGTGGTGTCAAAGGTGAAAGGC  
 AGTACTGGAACAGGACAGGAAGGATCTGAAACAAGTGGTGGTAAAGTGCCTGATAAAGG  
 TGTGTCTGTGAAGGAGTACCTGGTTGAGAAGTTTAAACCTCGTGAAGAAGATAAGGCGC  
 TTTCTGAGGTGATATCAGAGACTTTGCACAAGAAGAAAGAAGGGTCGGAAAAACAGGT  
 GAATCAAAGCCAATGGGGAAGGTGACTGAGTCAGAGGAGGTAGCAAGGCCTCTTGGGAC  
 TGGAATGGAGAGGAAGAGAGAGGATGAAGATGCCATTGCTGCTGGTCGCGAAAAGCAGTG  
 GAAAGGGTATGGTAGTAGGCTTAAAGGGGCTGTGAGCTCATGGATTGTTAAAGGTTCGT  
 GAAGACCAACAATATTTCTCAGGGGGCAGCTGATTGATCCTCAATGGTAAAAAATAAGCAAG  
 TTTTACCATTAAGAAATATGCATTCTATCCACCTTTTTCACTTCCACTGTTTCATGCTT  
 AATCCTAATAGGATTTTCCCATTTCTTTTGGCTCCTTGTACAGTAAGAAATGAAGGTTCTG  
 CAGCTAGTGATGAAATCGGGCATCGCAGACTCCAAGAATCAGGCAATTGATGCAAGTTG  
 GTCCTAGCTCTTCATCAGAGGTTTGGAGGCATGTTTCTGCTGCTGCTTTGCTACATGTT  
 TTTAGGAAATATGGTTCAAGACAAAGTAAAAAATACCTTTGTGTCTAGAGTTATATGGA  
 CTCTATGCTTATTATGTATAAAGGATGAATATAAATTAATGCTGCTGTGTTTGGATGTTGT  
 GCAAACATTGTCCGTTTGTATGAAGTTTGTGTTTGTAAAAAATAAATAAATAAATAA  
 AACCCGGAGGGGGGGCCGG

Legenda: Sequência nucleotídica do gene *RD29* de referência

A sequência (2149pb) referente ao Contig5590 contém os SNPs (SNP<sub>RD29-1</sub> até SNP<sub>RD29-6</sub>) que são identificados com cores diferentes (SNP<sub>RD29-1</sub> **A**, SNP<sub>RD29-2</sub> **G**, SNP<sub>RD29-3</sub> **A**, SNP<sub>RD29-4</sub> **C**, SNP<sub>RD29-5</sub> **G** e SNP<sub>RD29-6</sub> **T**). A sequência codante (111-1706) é indicada em negrito. Os *primers RD29-F* 5' TTTGTTGAAACAGGCATGGAATC3' e *RD29-R* 5' GAGCAAAGAATGGGAAAATCCT3' usados para a amplificação do gene estão sublinhados.

### Sequência da proteína *RD29* deduzida do Contig5590

MESQLHRPYDHTDDQDPQHAAVEDEGDHHHEKTSVIRKVKAKAKKIKETLAKHGLGYE  
 HEHERDYSHDDPDEGEENEDEEMEEDPEVHGAPMYESTVIGSAIPAQNVNLEKPTAIGE  
 DSYEGQRNAGKTGATRPALLEGQSGENLGKPSAMEVVHAPEDKDISFPPEIRRRTKGDSCGA  
 RNDNENVGPQGFRIGALEGLEEDPQAPKNRPGEVPPSNYQTKVTDPTGKGGEEVGITPL  
 IQSFNKMGVYDESVPKSESEQEAYTGSHGQFAPEPNPTEGKSDSAPKSCDPSKPEDNLP  
 RDTLTGKLSQSGYVEK**ISLATSTIADKAI**SAKNV**VASKLGYGG**TEGGKVPETDEKKNA  
 AKSGASPAEYAHKVSATVTDKLPVYQKVADAGSSVVSKVKGSTGTGQEGSETSGGKVP  
 DKGVSVKEYLVEKFKPREEDKALSEVISETLHKKKE**E**VGKTGESKPMGKVTESEEVAR**R**  
 LGTGMERKRE**E**DAIAAGRESSGKGM**V**RLKGAVSSWIVKGREDDQQYSQGAADSSNGKK

Legenda: Sequência da proteína putativa *RD29*

A sequência (531 aminoácidos) da proteína putativa *RD29* é apresentada junto com os aminoácidos relacionados com os SNP identificados: SNP<sub>RD29-1</sub> (**E**), SNP<sub>RD29-2</sub> (**R**), SNP<sub>RD29-3</sub> (**D**), SNP<sub>RD29-4</sub> (**E**), SNP<sub>RD29-5</sub> (**M**) e SNP<sub>RD29-6</sub> (**D**). O domínio CAP160 (pfam07918: aminoácidos I<sub>313</sub>-G<sub>339</sub>) é indicado em graxo.

## APÊNDICE D

### 1 Soluções para eletroforese em gel de agarose

#### **Tampão de corrida para gel de agarose (TAE) 50x:**

- a) 242,0 g Tris-base
- b) 100 ml 0,5 M EDTA
- c) Completa em 942,9 ml de água Milli-Q autoclavada.
- d) Adiciona 57,1 ml de Ácido Acético Glacial.

#### **Soluções**

##### **Para preparar 1 litro solução usa-se:**

- a) 0,5 M EDTA pH 8,0 (186,1g EDTA e 20g NaOH)
- b) Preparar 1 litro solução 1,0 M TrisHCl pH 7,5 (121,1g Tris ajustar pH 7,5 com HCl)

### 2 Soluções para isolamento de plasmídeo

#### **Solução I (GET)**

##### **Para preparar 500 ml de solução usa-se:**

- a) 23 ml de solução de glicose 20%
  - b) 10 ml de solução EDTA 0,5M a pH 8,0
  - c) 13 ml de solução de Tris 1M a pH 7,4
  - d) Completar o volume para 500 ml com água Milli-Q autoclavada
- Estocar aos 4 °C

#### **Solução II**

**Preparar somente na hora em que for usar. Para preparar 15 ml de solução usa-se:**

- a) 10 ml de água Milli-Q autoclavada
  - b) 750  $\mu$ L de solução de NaOH 4M
  - c) 1,5 ml de solução SDS 10%
  - d) Completar o volume para 15 ml com água Milli-Q autoclavada
- Não autoclavar a solução depois de pronta.

### **Solução III**

**Para preparar 100 ml de solução III usa-se:**

- a) 60 ml de solução de acetato de potássio 5M
- b) 11,5 ml de ácido acético glacial.
- c) 28,5 ml de água Milli-Q autoclavada (volume final de 100 ml)
- e) Estocar aos 4 °C

Não autoclavar a solução depois de pronta

### **3 Preparo de meio SOC**

Para preparar 100 ml de meio SOC usa-se.

- a) 02 g de Peptona
- b) 0,5 g de Yest Extract
- c) 1 ml de NaCl 1M
- d) 250  $\mu$ L de KCl 1M
- e) Completar o volume com água Milli-Q autoclavada para 98 ml
- f) Na capela adiciona 1 ml de  $Mg^{+2}$  1M (estéril) e 1 ml de Glucose 2M (filtrada)
- g) Alicotar em tubos falcon de 15 ml.

### **4 Preparo de meio LB (Luria-Bertani) para cultura de bactéria**

Para preparar 1 litro de meio LB usa-se:

- a) 10,0 g de NaCl
- b) 10,0 g de Peptona

- c) 05,0 g de Yest Extract
- d) Adicionar 800 ml de água Milli-Q e ajusta o pH 7,5 com NaOH 5M
- e) Completa o volume para 1 litro com água Milli-Q
- f) Autoclavar por 20 min.

OBS: Para preparar 1 litro de meio LB sólido acrescentar 7,5 g de Ágar bacteriológico para cada 500 ml de meio.