

# ANÁLISE *in silico* DAS BIBLIOTECAS DE cDNA SH2 E SH3 PARA A IDENTIFICAÇÃO DE GENES RESPONSIVOS À SECA EM CAFEIEIRO<sup>1</sup>

Felipe Vinecky<sup>2</sup>, Felipe Rodrigues da Silva<sup>3</sup>, Alan Carvalho Andrade<sup>4</sup>

(Recebido: 27 de outubro de 2009; aceito 19 de outubro de 2011)

**RESUMO:** O Brasil é o maior produtor e exportador de café no mundo e a cafeicultura é uma fonte de renda importante, para pequenos produtores. A seca, que vem se tornando cada vez mais intensa ao longo dos anos, prejudica a produção desses agricultores. Para auxiliar o desenvolvimento de plantas tolerantes à seca, vários grupos de pesquisa buscam uma melhor compreensão dos fatores genéticos envolvidos na resposta das plantas à seca. A construção e sequenciamento de bibliotecas ESTs (Expressed Sequence Tags) é um meio rápido e efetivo de se obter informações acerca da maioria dos genes expressos. O genoma funcional, do cafeeiro realizado por meio do sequenciamento de cDNAs (ESTs), possibilitou a construção de um amplo banco de dados de ESTs com seqüências de três espécies distintas de *Coffea*. A base de dados do genoma café constitui uma rica fonte de informações para estudos genéticos e fisiológicos do cafeeiro. Objetivou-se, no presente trabalho identificar genes candidatos (GC), potencialmente envolvidos na resposta ao estresse hídrico em cafeeiro, a partir de uma análise *in silico* dos ESTs disponíveis na base de dados do genoma café. Para essas análises foram utilizadas comparações *in silico* entre os dados de ESTs das bibliotecas SH2 (*Coffea arabica*) e SH3 (*Coffea canephora*), com o auxílio das ferramentas de bioinformática disponíveis na base de dados do genoma café. Com a metodologia utilizada, vários GC foram identificados e podem ser objeto de estudos experimentais posteriores, visando a seleção assistida por marcadores moleculares e para a rápida obtenção de variedades de café tolerantes à seca.

Palavras-chave: *Coffea arabica*, *Coffea canephora*, sequenciamento de cDNA, bioinformática, estresse hídrico.

## ***In silico* ANALYSIS OF cDNA LIBRARIES SH2 and SH3 FOR THE IDENTIFICATION OF GENES RESPONSIVE TO DROUGHT IN COFFEE**

**ABSTRACT:** Brazil is the largest producer and exporter of coffee and coffee production is a major source of income for small farmers. Drought, which has become increasingly intense over the years, affects the production of these farmers, as it reduces productivity and can even cause entire crop loss. Aiming at the development of drought tolerant crops, several research groups are currently studying the genetic factors involved in plant responses to drought. The construction and sequencing of EST libraries (Expressed Sequence Tags) is a quick and effective way to obtain information about most expressed genes. The functional genome of coffee made available by cDNA sequencing (ESTs), allowed the construction of a large EST database of three different *Coffea* species. The Coffee EST Database provides a rich source of information for genetic and physiological studies of coffee plants. The goal of this study was to identify genes potentially involved in the response of coffee plants to water stress by means of an (*in silico*) analysis of available ESTs of the coffee genome database. The methodology was based on *in silico* comparisons between EST data from libraries SH2 (*Coffea arabica*) and SH3 (*Coffea canephora*), with the aid of bioinformatics tools available at the Coffee Genome Database. With the methodology used, several candidate genes have been identified and may be subject to further experimental studies aiming at the establishment of a breeding program based on marker assisted selection for quick development of drought tolerant varieties of coffee.

Key words: *Coffea arabica*, *Coffea canephora*, cDNA sequencing, bioinformatics, drought stress.

## **1 INTRODUÇÃO**

Os projetos de genômica funcional são realizados a partir de seqüências expressas (ESTs), as quais são obtidas por meio do sequenciamento do DNA complementar (cDNA), derivado da transcrição

reversa do RNA mensageiro (mRNA) de um organismo, tecido ou ainda de um único tipo celular (VIEIRA et al., 2006). As ESTs são caracterizadas por serem sequenciadas uma única vez (single-pass) e por esse motivo, incompletas. Essas seqüências são geralmente pequenas, com cerca de 600 pares de

<sup>1</sup>Artigo extraído da Dissertação de Mestrado apresentada à Universidade Federal de Lavras/UFLA, pelo primeiro autor.

<sup>2</sup>Universidade de Brasília/UNB - Embrapa Recursos Genéticos e Biotecnologia - Parque Estação Biológica/PqEB - Av. W5 Norte (final) - Cx. P. 02372 - 70.770-917 - Brasília-DF fevinecky@yahoo.com.br

<sup>3</sup>Embrapa Informática Agropecuária - Laboratório de Bioinformática Aplicada Av. Dr. André Tosello -209 - Barão Geraldo, Cx. P. 06041 - 13.083-886 - Campinas-SP felipes@cnpia.embrapa.br

<sup>4</sup>Embrapa Recursos Genéticos e Biotecnologia - Laboratório de Genética Molecular (LGM) - Parque Estação Biológica - PqEB - Av. W5 Norte (final) - Cx. P. 02372 - 70.770-917 - Brasília-DF alan@cenargen.embrapa.br

bases (pb) e correspondem a sequências parciais do mRNA de origem. Uma vez sequenciadas as ESTs, torna-se possível a identificação de genes de interesse e também a caracterização do padrão de expressão gênica em diferentes condições ambientais.

Segundo Perret et al. (1998), a concentração de mRNA de um determinado gene em relação à concentração total de mRNA da célula é um indicativo do quanto esse gene está sendo expresso, isso é, do quanto a célula está investindo em seu maquinário bioquímico para produzir a proteína codificada pelo gene. Assim, foram desenvolvidas metodologias para medir a concentração relativa de determinados transcritos.

Vários organismos de interesse científico e/ou econômico têm sido investigados por meio de ESTs. Em tomate (*Solanum lycopersicum*), organismo utilizado como modelo para diversos estudos relacionado à biologia das plantas, foram obtidas mais de 150.000 ESTs a partir de 26 bibliotecas de cDNA (HOEVEN et al., 2002). Essas sequências estão sendo utilizadas para a identificação de genes específicos a tecidos e fases de desenvolvimento, e também para se identificar genes expressos em resposta aos estresses bióticos e abióticos (INSTITUTE FOR GENOMIC RESEARCH, 2001).

Com o avanço das tecnologias de sequenciamento de DNA, existem também bibliotecas de ESTs de trigo (ZHANG et al., 2004), milho (SHEN et al., 1994), arroz (YAMAMOTO; SASAKI, 1997), cana-de-açúcar (VETTORE et al., 2001), entre outras plantas. A maioria dos ESTs desses organismos encontram-se disponíveis nos bancos de dados publicados no National Center for Biotechnology Information - NCBI (1988).

O uso das ESTs tem facilitado o estudo de plantas como a *Medicago truncatula* e *Lotus japonicum* (HANDBERG; STOUGAARD, 1992) e permitiu, por exemplo, a identificação de vários genes com expressão diferencial em simbioses, como o NZP2213 (*Mesorhizobium loti* strain NZP2213 autoinducer synthase (mrlI2) gene). Para o estudo de genomas, com uma biblioteca de ESTs específica de nódulos de *M. truncatula*, 389 clones foram sequenciados, sendo que 34,2% das sequências não apresentaram similaridade significativa a sequências já depositadas em bancos de dados. Entre os clones que apresentaram similaridade significativa com

sequências já depositadas, foram encontradas 26 nodulinas, tais como *Early nodulin* (N-75), *Nodulin 25 precursor*, *Leghemoglobin 1*, dentre outras. Após a realização de *Northern blots* foi possível identificar genes relacionados à nodulação e genes específicos do nódulo (GYÖRGYÉY et al., 2000).

A construção de bibliotecas de ESTs é um meio rápido e efetivo de se produzir sequências parciais para a maioria dos genes expressos (NOGUEIRA et al., 2003). A comparação de uma sequência desconhecida com um vasto número de ESTs é um instrumento importante e poderoso da pesquisa genética (OHLROGGE; BENNING, 2000), e bancos de dados dessas sequências estão sendo utilizados na montagem de ferramentas, para a análise em larga escala da expressão gênica (CUSHMAN; BOHNERT, 2000).

O genoma funcional do cafeeiro foi realizado por meio do sequenciamento de cDNAs (ESTs), sintetizados a partir de material vegetal submetidos a diferentes tratamentos (VIEIRA et al., 2006), cada tratamento constituindo uma biblioteca, resultando na construção de um banco de dados com mais de 33.000 Unigenes de ESTs (*Expressed Sequence Tags* ou *Etiquetas de Sequências Expressas*) distintos de *C. arabica*, *C. canephora* e *C. racemosa*. A informação disponível nessa base de dados fornece uma rica fonte de informações para estudos da biologia e fisiologia das plantas de cafeeiro e certamente, facilitará a identificação e caracterização de genes de interesse agrônômico (VIEIRA et al., 2006).

Com a construção da base de dados torna-se possível separar grupos gênicos, potencialmente envolvidos na resposta do cafeeiro ao estresse hídrico, tais como fatores de transcrição, osmoprotetores, moléculas de proteção, transportadores de íons, bioenergética, metabolismo secundário, proteínas desconhecidas e/ou putativas, transdução de sinais, além do metabolismo de nucleotídeos, proteínas e lipídeos (INGRAM; BARTELS, 1996).

Na execução deste trabalho, foi realizada a identificação de genes envolvidos nas respostas do cafeeiro à seca, a partir da construção, sequenciamento e análise de bibliotecas de cDNA, oriundas de mRNA extraído de folhas coletadas de plantas de *C. arabica* (SH2-estresse hídrico em campo) e de um clone tolerante à seca de *C.*

*canephora* (SH3-estresse hídrico em vasos). Em estudos preliminares, foram identificados genes de café com alta similaridade aos envolvidos na resposta aos estresses abióticos, previamente descritos e caracterizados em outras espécies vegetais (VINECKY et al., 2005). A partir da expressão eletrônica (*in silico*), disponível na base de dados do genoma café, estudos comparativos dos resultados dos ESTs das bibliotecas SH2 e SH3, foram realizados neste trabalho. Essas análises permitiram a identificação de vários genes candidatos, com expressão “eletrônica” diferencial e associada à resposta ao estresse hídrico.

Para a confirmação prévia dos resultados obtidos nas análises *in silico*, foram realizados experimentos de PCR quantitativo em tempo real (qPCR), utilizando primers desenhados com base nas sequências dos genes identificados. Nesses experimentos de qPCR, foram utilizados clones de *C. canephora* que apresentam características diferenciadas para a tolerância à seca, tais como o crescimento radicular com maior profundidade, quando comparados a outros dois clones sensíveis ao estresse, onde as raízes foram encontradas mais próximas à superfície (PINHEIRO et al., 2005). A comparação da expressão de genes candidatos utilizando clones tolerantes e um sensível à seca, em que os clones de *C. canephora* com melhor rendimento e estabilidade durante a seca (clone 120), ou com capacidade de sobreviver a episódios de seca por meio de uma maior eficiência no uso da água (clone 14), podem fornecer informações importantes para uma melhor compreensão dos mecanismos fisiológicos e moleculares envolvidos na tolerância à seca do cafeeiro. Dessa forma, vários GC foram testados e serão publicados em outros trabalhos pelo grupo (MARRACCINI et al., 2009; VINECKY et al., 2009), mas o *contig* 18100 foi escolhido para este trabalho por codificar para o gene de um *no hit* e apresentar expressão diferencial em ambas as análises (*in silico* e qPCR).

## 2 MATERIAL E MÉTODOS

### 2.1 Material vegetal

Clones de *C. canephora* Pierre cv. Kouillou (também chamado de conilon no Brasil) tolerantes (clones 14, 73 e 120) e sensível (clone 22) à seca foram obtidos a partir de estacas enraizadas do Instituto de

Pesquisa e Assistência Rural (Incaper, Vitória, ES, Brasil) (FERRÃO; FONSECA; FERRÃO, 2000). Esses clones foram cultivados em casa de vegetação (25°C, 70% de umidade relativa, com média de fluxo de fótons fotossintético ao meio-dia de 900  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ), individualmente, em vasos de 12 litros com uma mistura de terra, areia e estrume de animais (3:1:1, v/v/v). Após 6 meses, as plantas de cada clone foram separadas em dois grupos: o primeiro recebeu uma irrigação regular (controle), enquanto a água foi suspensa durante a seca para o segundo (estresse hídrico). Para ambas as condições, 6 plantas (repetições biológicas) foram utilizadas. O estresse hídrico foi avaliado por meio da medição do potencial hídrico foliar de antemãhã ( $\Psi_{pd}$ ), com uma câmara de pressão tipo Scholander. O  $\Psi_{pd}$  foi regularmente acompanhado até atingir -3,0 MPa para plantas estressadas (SILVA, 2007). Tecidos frescos (folhas, 3 pares de ramos plagiotrópicos) foram coletados, sendo imediatamente congelados em nitrogênio líquido e armazenados a -80°C até serem utilizados para a extração de RNA e análises de expressão.

### 2.2 Extração de RNA

As amostras armazenadas a -80°C foram pulverizadas em nitrogênio líquido e o RNA total foi extraído usando o *Plant RNA Purification Reagent* (PRPR) (Invitrogen, São Paulo, Brasil). Cerca de 50mg de pó foi adicionado a 500 $\mu\text{L}$  de tampão PRPR, misturado vigorosamente a 25°C e, em seguida, centrifugado (16000 g, 2 min, 4°C). Após a adição de 5M NaCl (100 $\mu\text{L}$ ) e clorofórmio (300 $\mu\text{L}$ ) ao sobrenadante, a amostra foi centrifugada, como descrito anteriormente. Um volume de isopropanol foi ainda adicionado ao sobrenadante. Após incubação a 25°C durante 30min, os ácidos nucleicos foram precipitados por centrifugação (16000 g, 30 min, 4°C), o *pellet* foi seco e ressuspenso em 40 $\mu\text{L}$  de água livre de RNase e as amostras armazenadas a -20°C. A quantificação do RNA foi realizada, utilizando-se Espectrofotômetro (Waltham, MA, E.U.A.). Já a qualidade do RNA foi verificada por eletroforese em gel de agarose a 1,0% em 1X TAE e inspeção visual das bandas de RNA ribossomal, após coloração com 1% de brometo de etídio.

### 2.3 Purificação de mRNA

A purificação do RNA mensageiro (mRNA) foi realizada a partir do anelamento da cauda poli-A

com a cauda poli-T existente nas esferas de resina, presentes no Kit Oligotex® (Qiagen, Valencia, USA), de acordo com Ivanova et al. (2005).

#### 2.4 Construção das Bibliotecas de cDNA

A partir de 1-2 µg poli(A) + mRNA purificado dos tecidos de *C. arabica* (SH2) e *C. canephora* (SH3), a construção dessas bibliotecas foi feita seguindo o protocolo do *SuperScript II - Plasmid System and Plasmid Cloning Kit* (Invitrogen, São Paulo, Brasil), de acordo com Vieira et al. (2006). A eficiência da síntese de cDNA foi monitorada com nucleotídeos radioativos. Os cDNAs foram fracionados por diferença de tamanho utilizando-se uma coluna de Sepharose CL-2B. Frações contendo cDNA com mais de 500 pb foram ligadas em vetores pSPORT1 (Invitrogen), nos sítios de *Sall-NotI*. Os plasmídeos resultantes foram transformados por eletroporação em células de *E. coli* DH5α (Invitrogen).

O DNA plasmidial foi purificado utilizando um método modificado de lise alcalina (SAMBROOK; FRITSCH; MANIATIS, 1989). As reações de sequenciamento foram realizadas utilizando-se o ABI BigDye Terminator Sequencing kit (Applied Biosystems), as inserções de cDNA foram sequenciadas a partir da região 5' com o primer universal M13 REV, localizado no vetor pSPORT1 e os produtos da reação de sequenciamento foram analisados em um sequenciador ABI 3700 (Applied Biosystems), de acordo com Vieira et al. (2006).

#### 2.5 RT-PCR e PCR quantitativo em Real-time (qPCR)

Para eliminar a contaminação de DNA genômico, as amostras de RNA foram tratadas com RQ1 RNase-free DNase de acordo com as instruções do fabricante (Promega, Madison, WI, E.U.A.) e a qualidade do RNA foi verificada por eletroforese em gel e inspeção visual das bandas de RNA ribossomal, após coloração com 1% de brometo de etídio. A síntese da primeira fita de cDNA foi feita com o tratamento de 1mg de RNA total, utilizando-se o kit *ImProm-II™ Reverse Transcription System* com oligo (DT15), de acordo com as recomendações do fabricante (Promega). A ausência de contaminação do DNA genômico na preparação do cDNA foi verificada por PCR comum, utilizando pares de

primers SUS10 5'-GTTATCCTGATACCGGTG-3' e SUS11 5'-GGATCAAAAACATCAATGCC-3', que abrangem os íntrons 5 a 9 do gene CcSUS1 (AJ880768) que codifica a isoforma 1 da sacarose sintase (GEROMEL et al., 2006; LEROY et al., 2006). A RT-PCR foi realizada utilizando-se 1µL de DNA sintetizado sob a condição de PCR convencional, usando um termociclador PTC-100 (MJ Research) com GoTaq DNA polimerase de acordo com o fornecedor (Promega), nas seguintes condições: desnaturação inicial 94°C, 2 min, seguido de 40 ciclos de 94°C, 30 seg e 55°C, 30 seg e 72°C, 3 min e uma etapa de extensão final de 72°C, 6 min. Em tais condições, a amplificação de um fragmento de 667 bp caracteriza o cDNA CcSUS1 e a ausência de sequência genômica correspondente é indicada pela falta do fragmento de 1130pb, quando analisado em gel de agarose a 1% (dados não apresentados).

O qPCR foi realizado com cDNA fita simples, descrito acima, e utilizando o protocolo recomendado para o uso do 7500 Fast Real-Time PCR Systems (Applied Biosystems, Foster City, CA, E.U.A.). As amostras de cDNA foram diluídas (1/50 e 1/100) e testadas por qPCR usando pares de primers desenhados, utilizando-se o software *Primer Express* (Applied Biosystems) e testados preliminarmente para especificidade e eficiência utilizando um clone de EST (dados não apresentados). O qPCR foi realizado com 1 µL de cDNA e 0,2 µM (concentração final) de cada par de primers em um volume final de 10µL com fluorocromos SYBR green (SYBRGreen qPCR Mix-UDG/ROX, Invitrogen). A mistura de reação foi incubada por 2 min a 50°C (tratamento uracyl, Uracila-DNA), depois 5 minutos a 95°C (inativação de UDGase), seguido por 40 ciclos de amplificação de 3 seg. a 95°C, 30 seg. a 60°C. Os dados foram analisados usando o *SDS 2.1 software* (Applied Biosystems), para determinar o limiar de ciclo (Ct) valores. A eficiência do PCR (E) foi estimada usando dados de fluorescência absoluta obtidos durante a fase exponencial de amplificação de cada reação com a equação  $(1+E) = 10^{-1/\text{slope}}$  (RAMAKERS et al., 2003). Os níveis de expressão relativa do *contig18100* foram normalizados com os níveis de expressão do gene da ubiquitina usado como um controle interno (CRUZ et al., 2009) cujo par de primers foi BUBI F 5'-AAGACAGCTTCAACAGAGTACAGCAT-3' e

BUBI R 5'-GGCAGGACCTTGGCTGACTATA-3', e o par de primers do *contig* 18100 foi 5'-CTCGCGTGGCCGAGATT-3' e 5'-CCCTCACATTTCCACGTGAAT-3'. A especificidade dos produtos de PCR gerados para cada conjunto de primers foi verificada pela análise da Tm (dissociação) dos produtos amplificados.

## 2.6 Tratamento dos dados em bioinformática

Na Embrapa, uma base de dados relacional foi desenvolvida, a fim de anotar e acessar prontamente a informação, utilizando-se o sistema MySQL (MY STRUCTURED QUERY LANGUAGE, 1995). A partir da submissão das sequências para a Base de Dados (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA, 2006), foram realizadas as seguintes etapas, seguindo os procedimentos de Telles e Silva (2001): 1. Análise da qualidade das sequências; 2. Limpeza (Eliminação de sequências ribossomais, vetores, baixa qualidade, tamanho reduzido, etc); 3. *Clusterização* das sequências; 4. Análises de BlastX contra a base de dados não-redundantes (NR) do NCBI (National Center for Biotechnology Information); 5. Armazenamento da informação. A qualidade das sequências foi definida com os parâmetros de tamanho sendo > 350 pb e de Phred  $\geq 20$ .

As análises *in silico* comparativas deste trabalho, foram feitas a partir da *clusterização* e montagem, disponíveis na Base de Dados do Genoma Café (EMBRAPA, 2006). Os dados foram obtidos utilizando-se as ferramentas de listar bibliotecas disponíveis e consulta por biblioteca específica, utilizando-se SH2, SH3 e SH2 SH3 como palavras de procura. Para os cálculos de expressão relativa foram utilizados os valores de ocorrência de ESTs (SH2 ou SH3) para cada *cluster*, normalizado pelo número de ESTs válidos para cada biblioteca, multiplicado por 100. Para a biblioteca SH2, o valor total de ESTs válidos foi de 5053 e para SH3 foi 5743. A expressão relativa ao *cluster* mais expresso de cada biblioteca foi calculado, multiplicando-se o resultado anterior pelo fator 74,30882 e para SH3 pelo fator 37,535795. Esses valores foram calculados com base nos resultados dos *clusters* com maior ocorrência de ESTs das bibliotecas SH2 e SH3, respectivamente. Para SH2, o *cluster* mais expresso

foi o *contig* 10569 - rubisco small subunit - *Coffea arabica* descrito por Marraccini et al. (2004), com 68 ESTs de SH2, sendo o valor de 1,34 e portanto,  $100/1,34 = 74,30882$  e para SH3 o *contig* mais expresso foi o *contig* 15182 (oxidoreductase - *Arabidopsis thaliana*), sendo o valor de expressão normalizado de 2,66 e portanto,  $100/2,66 = 37,535795$ .

## 3 RESULTADOS E DISCUSSÃO

### 3.1 Caracterização das bibliotecas de cDNA de café (SH2 e SH3).

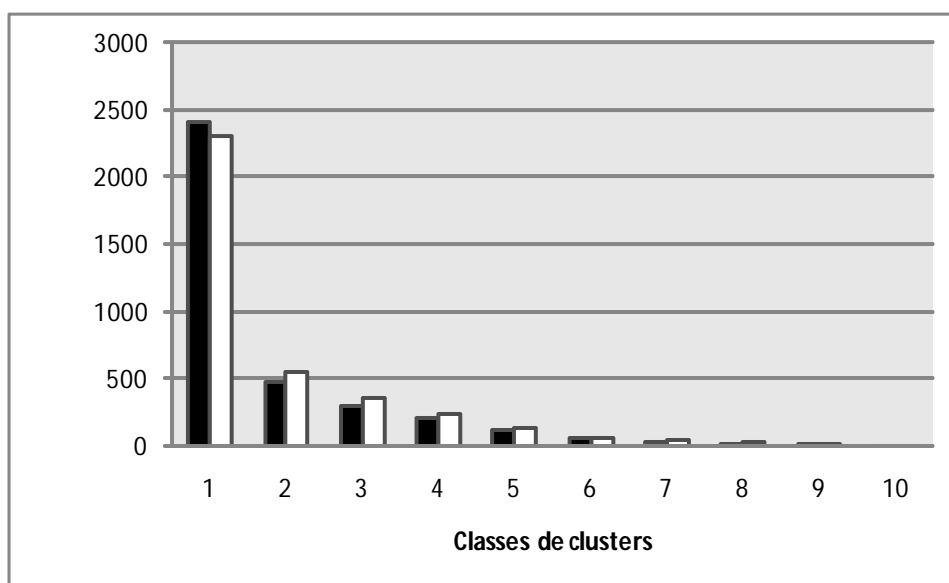
Um sumário dos dados de ESTs produzidos pelas bibliotecas de cDNA denominadas SH2 e SH3, construídas no âmbito do Projeto Genoma Café, a partir de material vegetal submetido ao estresse hídrico é apresentado na Tabela 1. Esses dados foram obtidos após o processamento de bioinformática das sequências EST de café e mostram que a partir de aproximadamente 15 mil sequências gênicas produzidas por clones dessas duas bibliotecas, foi disponibilizado para análises posteriores, um conjunto de mais de 10 mil sequências gênicas de boa qualidade e potencialmente envolvidas na resposta ao estresse hídrico em cafeeiro. Além disto, o índice de novidade de 66,5 % (SH2) e 59,1 % (SH3) das sequências destas bibliotecas foi muito bom em se tratando de bibliotecas de folhas, em que redundância é geralmente alta. A distribuição das sequências nos *contigs* também evidencia a baixa redundância, pois a maioria dos *contigs* é formada por poucas sequências (Figura 1).

### 3.2 Análises dos resultados de BLASTX dos *clusters* formados com sequências de SH2 e SH3.

Foi possível observar que, dos 3.360 *clusters* formados com as ESTs provenientes da biblioteca SH2, 2675 são *contigs* e 685, *singlets* (tabela 2). O resultado das análises de BLASTX, indicam que, no caso dos *contigs*, 2.449 apresentaram similaridade com alguma sequência presente na Base de Dados do NCBI (e-value  $\leq 10^{-5}$ ) e para 226 *contigs* a identificação dessa similaridade não foi possível. Já no caso dos *singlets*, 379 apresentaram resultado significativo de BLASTX, com a identificação de similaridade (e-value  $\leq 10^{-5}$ ) e 306 não apresentaram resultado significativo.

**TABELA 1** – Resumo dos dados de ESTs produzidos a partir de material vegetal de café, em condições de estresse hídrico.

Bib.de cDNA	Total de seqs.	Tamanho $\geq$ 300 pb Phred $>$ 20	Seqs. após trimagem	Índice de novidade(%)	Clusters
SH2	6824	5185 (75,98%)	5053	66,5	3360
SH3	7368	5886 (79,88%)	5743	59,1	3393



**FIGURA 1** – Distribuição dos *clusters* (*Contigs*) das bibliotecas de cDNA, SH2 (preto) e SH3 (branco), em função do número de sequências presentes em cada *cluster*.

**TABELA 2** – Síntese dos resultados de BLASTX (e-value  $\leq 10^{-5}$ ) dos *clusters*, formados com sequências de SH2 e SH3

	SH2	SH3	SH2 SH3	Unigene Seca
<i>Cluster</i>	3360	3393	-	6083
<i>Contigs</i>	2675	2625	670	4630
<i>Contigs com Blast hit</i>	2449	2421	642	4228
<i>Contigs No hit</i>	226	204	28	402
<i>Singlets</i>	685	768	-	1453
<i>Singlets com Blast hit</i>	379	472	-	851
<i>Singlets No hit</i>	306	296	-	602

Observa-se também que as ESTs da biblioteca SH3, foram agrupadas em 2625 *contigs* e 768 *singlets*, perfazendo assim um total de 3393 *clusters*. Os resultados de BLASTX com essas sequências

indicaram que 2421 *contigs* e 472 *singlets* apresentaram resultado significativo de BLASTX (e-value  $\leq 10^{-5}$ ), sendo que, para 204 *contigs* e 296 *singlets* os resultados de BLASTX não foram significativos. A

comparação dos *clusters* de SH2 e SH3, indica que 670 *contigs* são comuns, ou seja, são formados por ESTs tanto de SH2 como de SH3, e desses, 642 possuem resultado significativo de BLASTX e 28, não. Dessa forma, o conjunto não redundante de ESTs provenientes de condições de estresse hídrico é de 4630 *contigs* e 1453 *singlets*, totalizando 6083 *clusters* provenientes da montagem cafEST.

### 3.3 Análise dos resultados de BLASTX dos *clusters* de SH2 e SH3

Considerando-se apenas os *contigs* com a ocorrência de mais de 4 ESTs, foi possível identificar 92 e 117 *contigs* provenientes das bibliotecas SH2 e SH3, respectivamente. Esse resultado indica que a maior parte dos *contigs* das bibliotecas SH2 e SH3, são formados com poucos (1 a 4) ESTs dessas bibliotecas. Isso pode ser verificado também, analisando-se a distribuição dos ESTs na Figura 1. Desse total de 209 *contigs*, com mais de 4 ESTs das bibliotecas SH2 e SH3, verifica-se que 26 são comuns e ocorrem nas duas bibliotecas (SH2 e SH3) e 183 são exclusivos e apresentam somente ESTs de SH2 ou SH3. A expressão relativa desses 26 *contigs* nas bibliotecas SH2 e SH3 (Figura 2) e o melhor resultado de BLASTX, correspondente, para cada um desses 26 *contigs* (Tabela 3), estão apresentados. Pode-se observar nessa lista a ocorrência de genes já caracterizados como envolvidos na resposta ao estresse hídrico, em outras espécies vegetais. É o caso do *contig* 9073 que codifica o gene da catalase que é uma enzima sempre utilizada pela planta para atuar contra o estresse oxidativo. Ela pode atuar detoxificando peróxido de hidrogênio e também no ciclo fotossintético, retirando espécies reativas de oxigênio, sendo expressa tanto em condições normais como em situações de estresse (INZE; MONTAGU, 1995). E também da proteína CDSP32 representada pelo *contig* 18601, que é uma proteína cloroplasmática induzida durante o estresse hídrico e que foi previamente identificada em plantas *Solanum tuberosum* submetidas ao déficit hídrico (EYMERY; REY, 1999).

Analisando a expressão relativa *in silico* dos 20 *contigs* mais expressos das bibliotecas SH2 e SH3 (Figura 3) com o melhor resultado de BLASTX correspondente a cada um desses *contigs* (Tabela 4),

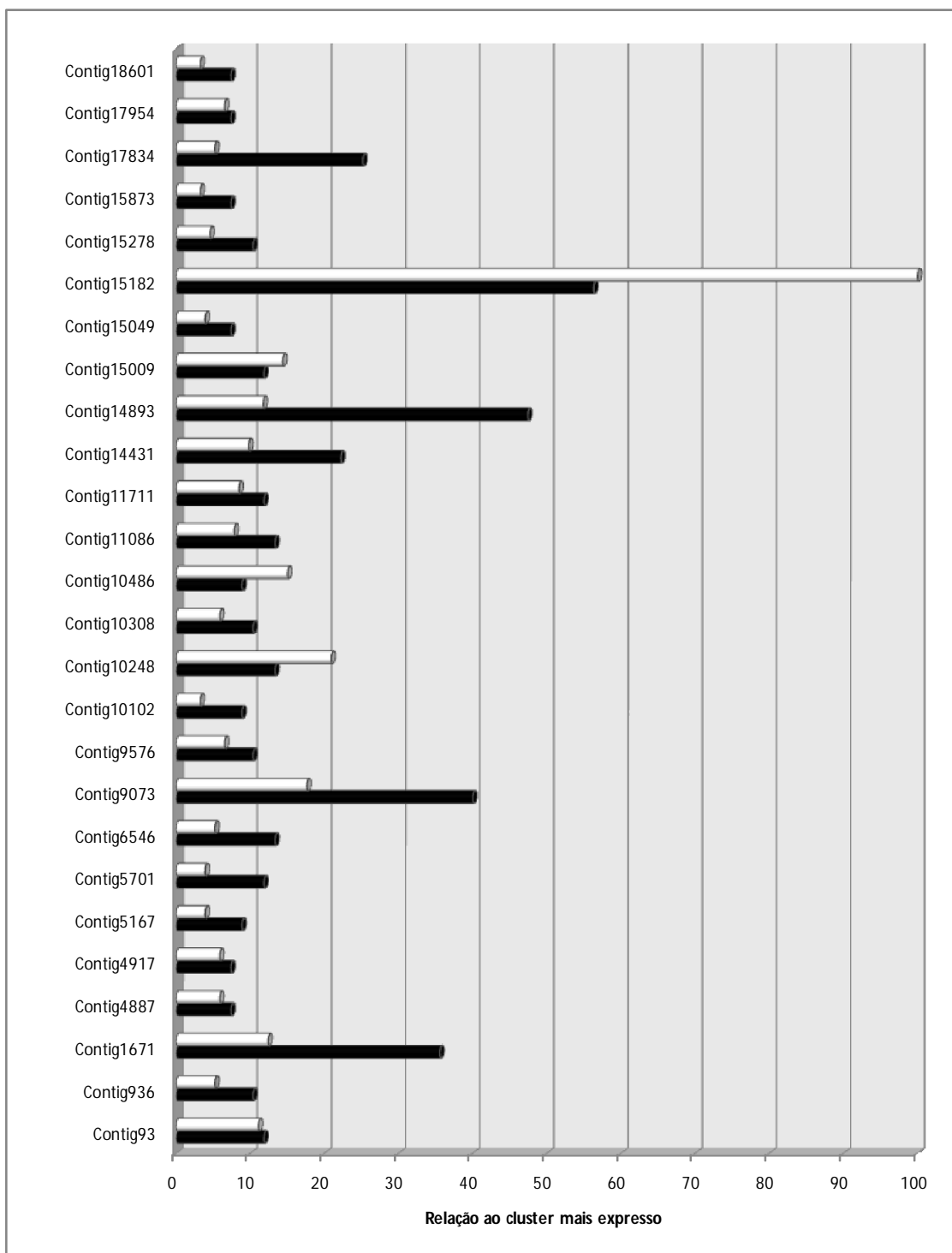
os dados indicam que o *contig* 10569 é o que apresenta maior expressão *in silico*, nessa biblioteca.

O *contig*10569 indica similaridade com a Rubisco, resultado esperado, uma vez que a biblioteca SH2 foi construída a partir de RNA total extraído de folhas. O gene da ribulose 1,5 bifosfato (Rubisco) é altamente expresso em plantas inclusive em *C. arabica* (MARRACCINI et al., 2003), estando entre os primeiros genes clonados e analisados em plantas, sendo que a sua regulação também tem sido extensivamente estudada. Gao e Loescher (2000) identificaram uma associação entre as atividades do Gliceraldeído-3-Fosfato Desidrogenase (G3PDH) e a mannose-6-fosfato redutase com os níveis de expressão do mRNA em resposta ao estresse salino e também durante o desenvolvimento foliar. E esse gene está representado, nas análises, pelo *contig* 15182, e que, no caso da biblioteca SH3, foi o que apresentou maior expressão *in silico*, sendo o *contig* com similaridade com a rubisco o 3° mais expresso, nessa biblioteca. O dado pode ter significado biológico, uma vez que a biblioteca SH3 foi construída a partir de material vegetal de um clone de conillon, previamente caracterizado como tolerante à seca (DAMATTA et al., 2002). Além disso, os dados de expressão *in silico*, na biblioteca SH2 (3° mais expresso) é mais um indicativo de que o *contig* 15182 pode estar envolvido na resposta ao estresse hídrico em café. Por último, a diferença dos *contigs* com similaridade à rubisco, pode ser também explicada pela diferença dos materiais genéticos utilizados para a construção das bibliotecas.

Além do *contig* 15182, mencionado acima, 4 outros (*contigs*: 1671, 9073, 14431 e 14893) aparecem na lista dos 20 genes mais expressos, tanto nas bibliotecas de SH2 e SH3 e são potenciais candidatos para o envolvimento na resposta ao estresse hídrico em cafeeiro (Figura 3). A presença de um *contig* com similaridade a uma catalase (*contig* 9073), nessa lista, corrobora com o potencial envolvimento desses genes na proteção ao mecanismo oxidativo causado pelo estresse hídrico (INZE; MONTAGU, 1995).

### 3.4 Análise dos *clusters* formados exclusivamente com ESTs de SH2, SH3 e SH2 SH3.

Os resultados de BLASTX dos 20 *clusters* com maior expressão *in silico* e formados



**FIGURA 2** – Dados de expressão relativa ao *cluster* mais expresso de cada biblioteca, para os 26 *clusters* comuns às bibliotecas SH2 (preto) e SH3 (branco).

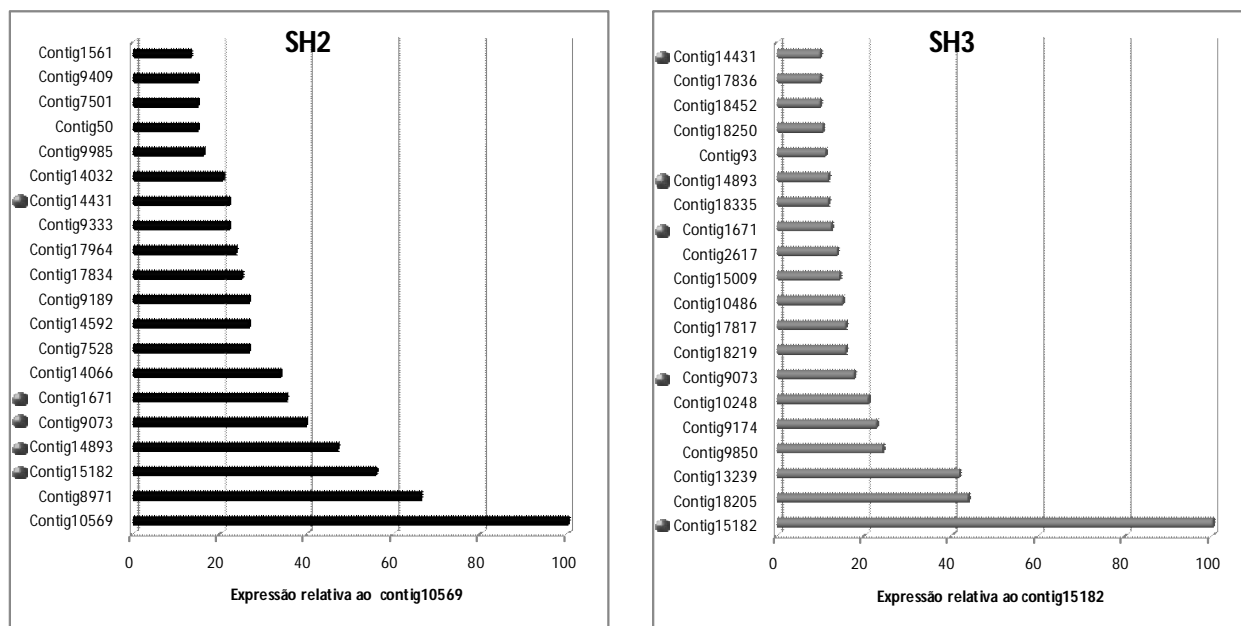


**TABELA 3** – Resultados de BLASTX dos 26 *clusters* comuns às bibliotecas SH2 e SH3, com expressão *in silico* maior que 4 ESTs.

<i>Contig</i>	<b>Blast Hit   Função [Espécie em que foi identificada]</b>
18601	emb CAA71103.1  CDSP32 protein [ <i>Solanum tuberosum</i> ]
17954	gb AAA34075.1  glyceraldehyde-3-phosphate dehydrogenase A-subunit prec. [ <i>Nicotiana</i> ]
17834	gb ABE83482.1  AAA ATPase, central region; Homeodomain-like [ <i>Medicago truncatula</i> ]
15873	ref NP_567574.1  ATP binding / kinase/ protein kinase... [ <i>Arabidopsis thaliana</i> ]
15278	gb ABA96025.1  flavin-containing monooxygenase, putative [ <i>Oryza sativa</i> ]
15182	ref NP_179721.1  oxidoreductase [ <i>Arabidopsis thaliana</i> ]
15049	gb ABE93728.1  Ferric reductase-like transmembrane component [ <i>Medicago truncatula</i> ]
15009	emb CAB51533.1  galactinol synthase, isoform GolS-1 [ <i>Ajuga reptans</i> ]
14893	gb ABE83482.1  AAA ATPase, central region; Homeodomain-like [ <i>Medicago truncatula</i> ]
14431	emb CAA36380.1  unnamed protein product [ <i>Gossypium hirsutum</i> ]
11711	emb CAB16918.1  P-Protein precursor [ <i>Solanum tuberosum</i> ]
11086	dbj BAA03131.1  glycolate oxidase [ <i>Cucurbita</i> cv. Kurokawa Amakuri]
10486	gb ABC59515.1  chloroplast photosystem II 22 kDa component [ <i>Nicotiana benthamiana</i> ]
10308	gb AAM22683.1  carbonic anhydrase [ <i>Gossypium hirsutum</i> ]
10248	gb AAA18529.1  chlorophyll A/B binding protein
10102	gb AAA34140.1  chlorophyll a/b-binding protein prf]
9576	dbj BAD45891.1  polyubiquitin [ <i>Oryza sativa</i> (japonica cultivar-group)]
9073	emb CAA85426.1  catalase [ <i>Nicotiana plumbaginifolia</i> ]
6546	emb CAA62364.1  L1 protein [ <i>Arabidopsis thaliana</i> ]
5701	gb ABG73417.1  chloroplast pigment-binding protein CP26 [ <i>Nicotiana tabacum</i> ]
5167	gb AAD27877.1  LHCII type III chlorophyll a/b binding protein [ <i>Vigna radiata</i> ]
4917	gb AAY89374.1  beta-amylase 1 [ <i>Nicotiana langsdorffii</i> x <i>Nicotiana sanderae</i> ]
4887	emb CAA66667.1  polyubiquitin [ <i>Pinus sylvestris</i> ]
1671	No hit
936	gb ABA86964.1  glyceraldehyde-3-phosphate dehydrogenase B subunit [ <i>Glycine max</i> ]
93	gb AAQ56195.1  aminotransferase 2 [ <i>Cucumis melo</i> ]

exclusivamente com ESTs das bibliotecas SH2 ou SH3 foram listados na Tabela 5. Quanto aos resultados de BLASTX correspondentes aos *clusters* formados exclusivamente com ESTs das bibliotecas SH2 e SH3, estão listados na tabela 6. Todos esses *clusters* foram identificados por meio

da ferramenta de busca de “Biblioteca Específica”, disponível na Base de Dados caEST. Foi analisada a expressão relativa *in silico* dos 20 *clusters* com maiores valores, obtidos da lista de *contigs* formados exclusivamente com ESTs de SH2 ou SH3 (Figura 4). Importante notar que três dos



**FIGURA 3** – Dados de expressão relativa ao *cluster* mais expresso de cada biblioteca, para os 20 *clusters* com maior expressão *in silico* das bibliotecas SH2 e SH3. As esferas ilustram os *clusters* comuns às duas bibliotecas.

*clusters* formados exclusivamente por ESTs, da biblioteca SH3 (Figura 4) também estão listados na Figura 3, indicando que, além de serem formados exclusivamente por ESTs dessa biblioteca, também figuram entre os 20 *clusters* com maior expressão *in silico*, da biblioteca SH3.

Observando-se os dados dos *contigs* 14838, 15284 (descritos abaixo) e 18002 - catalase descrito por Inze e Montagu (1995), apresentados na Figura 5, nota-se que esses apresentam uma maior ocorrência de ESTs provenientes da biblioteca SH3 (clone tolerante), quando comparado aos demais *clusters* formados exclusivamente por ESTs, das bibliotecas SH2 e SH3. Com os resultados de BLASTX correspondentes a esses *contigs* (Tabela 6) pode-se salientar que o *contig* 18002, apresenta similaridade com uma catalase. Nota-se também, que dentre os 10 *contigs* com maior expressão *in silico* (Figura 5), os *contigs* 18283 (unknown protein), 15208 (putative protein), 18467 (heat shock proteins - hsp), 18015 (cytosolic class II low molecular weight - hsp) e 15055 (no hit) são os que apresentam uma maior ocorrência de ESTs provenientes da biblioteca SH2 (estresse hídrico em campo). Dois desses *contigs* (18467 e 18015)

apresentam similaridade com heat shock proteins, já caracterizadas em outras espécies vegetais como componentes da resposta ao estresse hídrico, pois essas proteínas formam um sistema conservado de resposta celular para o mecanismo de defesa, que é ativado por uma variedade de estímulos tais como temperatura elevada e certa concentração de componentes químicos. É caracterizada por uma rápida reprogramação da expressão gênica, levando ao acúmulo de proteínas de choque térmico (hsp), a maioria das quais atuam como chaperonas (MORIMOTO, 1998; SCHÖFFL; PRÄNDL; REINDL, 1998).

O *contig* 14838 (Figura 5) demonstra certa similaridade ao gene da carbonate desidratase/zinc ion binding, que é um gene normalmente expresso, pois está presente no metabolismo de proteínas (CHO et al., 2006), porém também serve como um regulador da homeostase na planta. Já o *contig* 15284 codifica uma TPA (isoflavone redutase), e esse gene foi encontrado expresso também em biblioteca de ESTs de Trigo (*Triticum* ssp.), quando analisada a genômica funcional sob estresse abiótico (HOUDE et al., 2006).

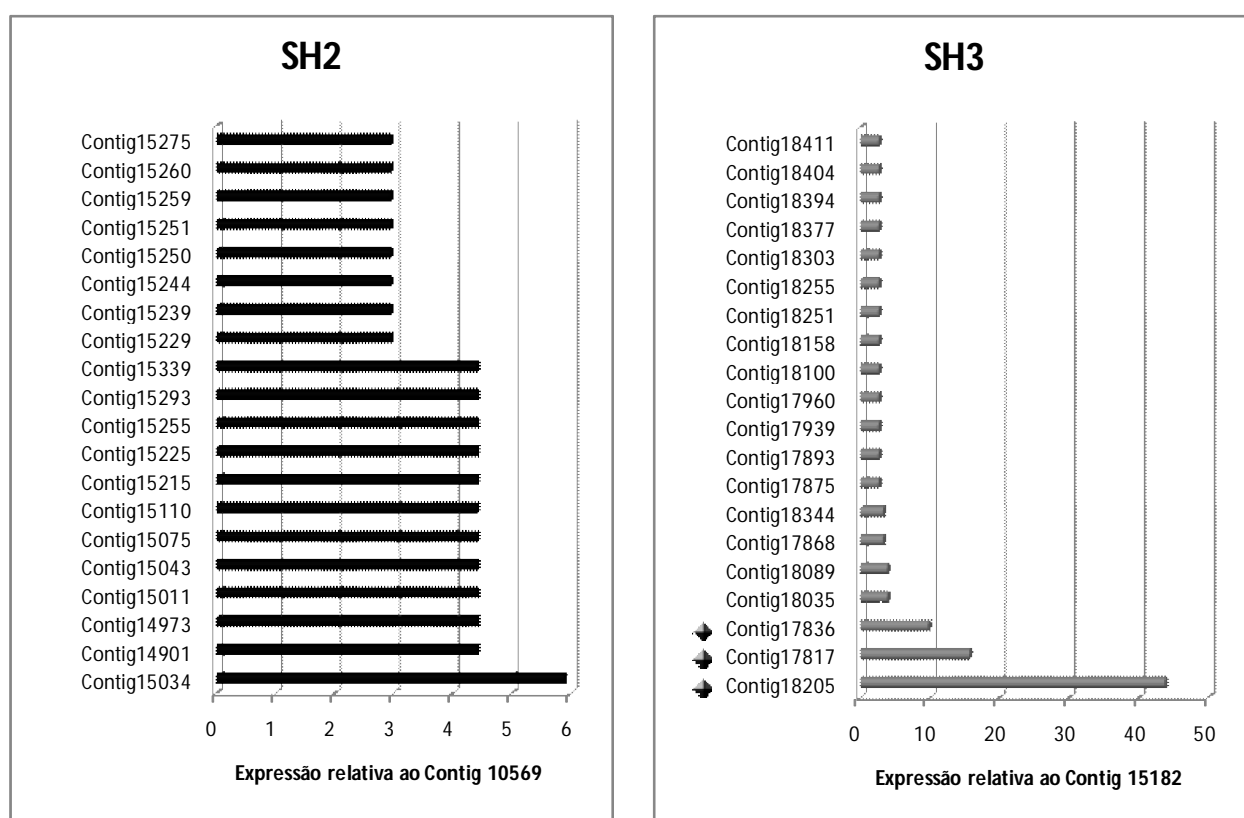
**TABELA 4** – Resultados de BLASTX dos 20 clusters com maior expressão *in silico* nas bibliotecas SH2 (superior) e SH3 (inferior). Os asteriscos apontam os clusters comuns às duas bibliotecas (SH2 e SH3).

SH	Contig	Blast Hit   Função [Espécie em que foi identificada]
2	Contig1561	gb ABE87096.1  Glycosidehydrolase, family 18 [Medicago truncatula]
2	Contig9409	dbj BAA89317.1  23kDa polypept. of the oxygen comp. photos. II [Cucumis sativus]
2	Contig7501	gb AAL01886.1  chitinase 3-like protein precursor [Trichosanthes kirilowii]
2	Contig50	gb AAB31705.1  photosystem I subunit PSI-E [Nicotiana sylvestris]
2	Contig9985	gb ABB29942.1  S-adenosyl methionine synthase-like [Solanum tuberosum]
2	Contig14032	gb AAO33591.1  putative early light induced protein [Arachis hypogaea]
2	*Contig14431	emb CAA36380.1  unnamed protein product [Gossypium hirsutum]
2	Contig9333	gb ABE83482.1  AAA ATPase, central region; Homeodomain [Medicago truncatula]
2	Contig17964	ref NP_179721.1  oxidoreductase [Arabidopsis thaliana]
2	Contig17834	gb ABE83482.1  AAA ATPase, central region; Homeodomain [Medicago truncatula]
2	Contig9189	No hit
2	Contig14592	gb AAL01886.1  chitinase 3-like protein precursor [Trichosanthes kirilowii]
2	Contig7528	gb ABE79940.1  Aldo/ketoreductase [Medicago truncatula]
2	Contig14066	ref NP_199832.1  quinolinate synthetase A [Arabidopsis thaliana]
2	*Contig1671	No hit
2	*Contig9073	emb CAA85426.1  catalase [Nicotiana glauca]
2	*Contig14893	gb ABE83482.1  AAA ATPase, central region; Homeodomain [Medicago truncatula]
2	*Contig15182	ref NP_179721.1  oxidoreductase [Arabidopsis thaliana]
2	Contig8971	gb AAL01886.1  chitinase 3-like protein precursor [Trichosanthes kirilowii]
2	Contig10569	emb CAD11991.1  rubisco small subunit [Coffea arabica]
3	*Contig14431	emb CAA36380.1  unnamed protein product [Gossypium hirsutum]
3	Contig17836	gb AAO33591.1  putative early light induced protein [Arachis hypogaea]
3	Contig18452	gb AAD27882.2  chlorophyll a/b-binding protein CP24 precursor [Vigna radiata]
3	Contig18250	No hit
3	Contig93	gb AAQ56195.1  aminotransferase 2 [Cucumis melo]
3	*Contig14893	gb ABE83482.1  AAA ATPase, central region; Homeodomain [Medicago truncatula]
3	Contig18335	No hit
3	*Contig1671	No hit
3	Contig2617	ref NP_179721.1  oxidoreductase [Arabidopsis thaliana]
3	Contig15009	emb CAB51533.1  galactinol synthase, isoform GOLS-1 [Ajuga reptans]
3	Contig10486	gb ABC59515.1  chloroplast photos. II 22 kDa component [Nicotiana glauca]
3	Contig17817	gb AAO33591.1  putative early light induced protein [Arachis hypogaea]
3	Contig18219	No hit

Continua...

TABELA 4 – Continuação...

SH	Contig	Blast Hit   Função [Espécie em que foi identificada]
3	*Contig9073	emb CAA85426.1  catalase [Nicotianaplumbaginifolia]
3	Contig10248	gb AAA18529.1  chlorophyll A/B binding protein[Arabidopsis thaliana]
3	Contig9174	emb CAE05952.3  OSJNBb0088C09.11 [Oryza sativa (japonica cultivar-group)]
3	Contig9850	emb CAA88629.1  pre-pro-cysteine proteinase [Lycopersiconesulentum]
3	Contig13239	emb CAD11991.1  rubisco small subunit [Coffeearabica]
3	Contig18205	No hit
3	*Contig15182	ref NP_179721.1  oxidoreductase [Arabidopsisthaliana]



**FIGURA 4** – Dados de expressão relativa ao *cluster* mais expresso de cada biblioteca, para os 20 *clusters* com maior expressão *in silico*, formados exclusivamente com ESTs das bibliotecas SH2 e SH3. As pirâmides ilustram os *clusters* que também estão entre os 20 *contigs* com maior expressão *in silico*, da biblioteca SH3.

**TABELA 5** – Resultados de BLASTX dos 20 *clusters* com maior expressão *in silico*, formados exclusivamente com ESTs das bibliotecas SH2 e SH3.

SH	Contig	Blast Hit   Função [Espécie em que foi identificada]
2	Contig15229	No hit
2	Contig15239	gb AAU94373.1  At2g36070 [ <i>Arabidopsis thaliana</i> ]
2	Contig15244	gb AAS49088.1  At2g24395 [ <i>Arabidopsis thaliana</i> ]
2	Contig15250	No hit
2	Contig15251	ref YP_545679.1  methyl-accep. chemotaxis sens. transd. [ <i>Methylobacillus flagellatus</i> ]
2	Contig15259	No hit
2	Contig15260	No hit
2	Contig15275	ref XP_482757.1  putative elicitor-inducible cytochrome P450 [ <i>Oryza sativa</i> ]
2	Contig14901	No hit
2	Contig14973	ref YP_532907.1  ABC transporter related [ <i>Rhodopseudomonas palustris BisB18</i> ]
2	Contig15011	ref ZP_00848179.1  ABC transporter, related [ <i>Rhodopseudomonas palustris BisB18</i> ]
2	Contig15043	ref NP_197514.2  binding [ <i>Arabidopsis thaliana</i> ]
2	Contig15075	dbj BAA88921.1  sigma factor [ <i>Nicotiana tabacum</i> ]
2	Contig15110	gb AAB03097.1  Hsp22.3
2	Contig15215	No hit
2	Contig15225	No hit
2	Contig15255	emb CAB79380.1  hypothetical protein [ <i>Arabidopsis thaliana</i> ]
2	Contig15293	gb AAO22570.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
2	Contig15339	No hit
2	Contig15034	ref NP_850380.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
3	Contig18411	ref XP_463936.1  putative NADPH [ <i>Oryza sativa</i> (japonica cultivar-group)]
3	Contig18404	No hit
3	Contig18394	dbj BAA96917.1  unnamed protein product [ <i>Arabidopsis thaliana</i> ]
3	Contig18377	gb AAF86687.1  MTD1 [ <i>Medicago truncatula</i> ]
3	Contig18303	emb CAB87849.1  receptor kinase-like protein [ <i>Arabidopsis thaliana</i> ]
3	Contig18255	dbj BAF08985.1  Os02g0543000 [ <i>Oryza sativa</i> (japonica cultivar-group)]
3	Contig18251	dbj BAC77694.1  lipid transfer protein [ <i>Atriplex nummularia</i> ]
3	Contig18158	gb AAZ32409.1  S-methyltransferase [ <i>Catharanthus roseus</i> ]
3	Contig18100	No hit

Continua...

TABELA 5 – Continuação...

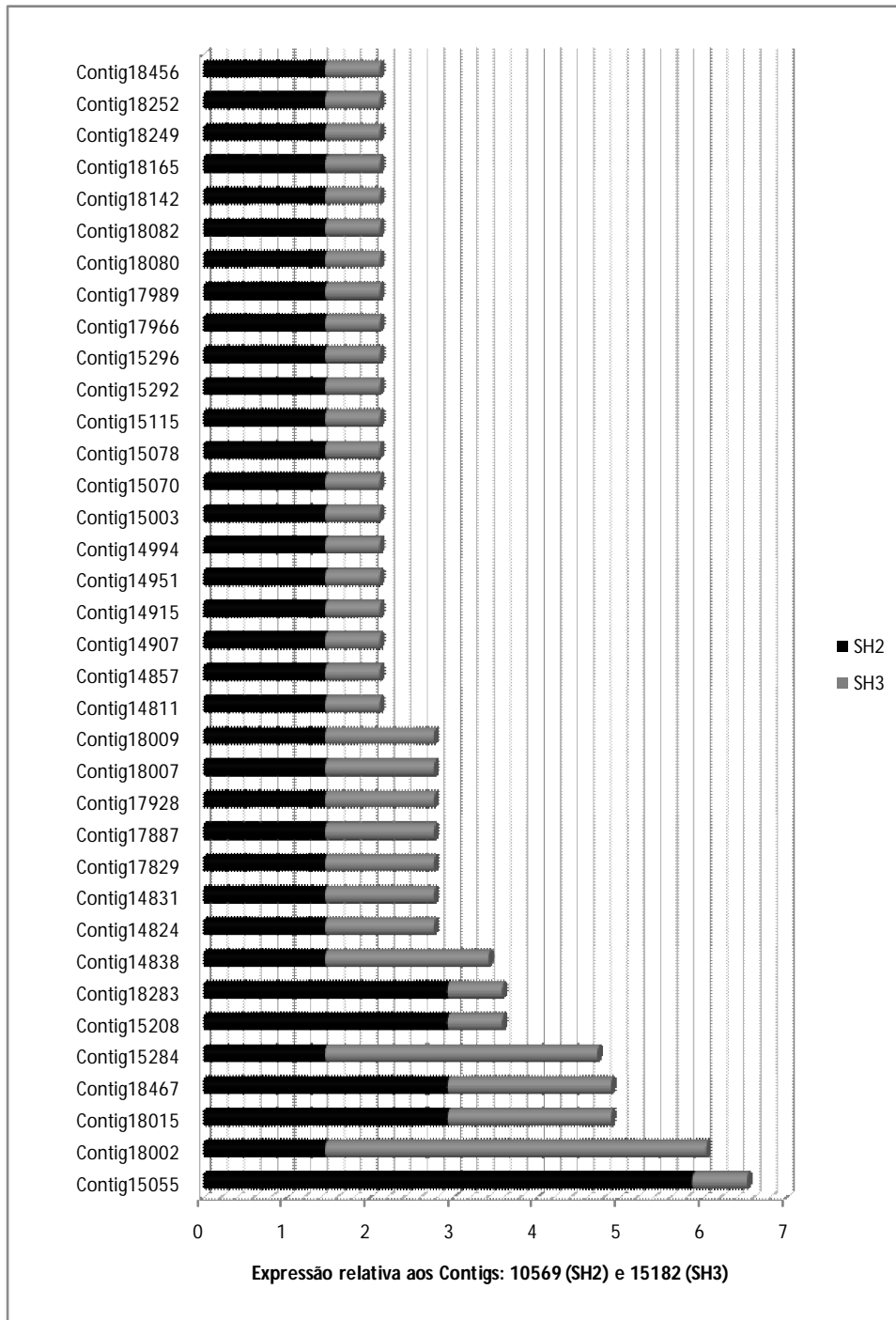
SH	Contig	Blast Hit   Função [Espécie em que foi identificada]
3	Contig17960	No hit
3	Contig17939	No hit
3	Contig17893	ref NP_564667.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
3	Contig17875	emb CAA77213.1  sigma-like factor [ <i>Arabidopsis thaliana</i> ]
3	Contig18344	gb AAC24001.1  isoflavone reductase related protein [ <i>Pyrus communis</i> ]
3	Contig17868	gb ABE83633.1  Ribosomal protein L10; Rib. protein 60S [ <i>Medicago truncatula</i> ]
3	Contig18089	gb AAP85545.1  putative WRKY-type DNA binding protein [ <i>Glycine max</i> ]
3	Contig18035	gb AAO85557.1  photosystem I subunit XI [ <i>Nicotiana attenuata</i> ]
3	Contig17836	gb AAO33591.1  putative early light induced protein [ <i>Arachis hypogaea</i> ]
3	Contig17817	gb AAO33591.1  putative early light induced protein [ <i>Arachis hypogaea</i> ]
3	Contig18205	No hit

### 3.5 Análise de PCR quantitativo em tempo real (qPCR)

Vários pares de primers foram utilizados para a realização das reações de qPCR com a finalidade de comprovar, experimentalmente, a expressão de genes candidatos identificados por meio das análises *in silico*. Resultados parciais desses experimentos já foram publicados em trabalhos anteriores (ALVES et al., 2008; VINECKY et al., 2008a, 2008b, 2009). Após a identificação *in silico* e experimentos de macroarranjo (VINECKY et al., 2009), o *contig* 18100 (*No hit*), foi selecionado para a análise de qPCR deste trabalho. A expressão gênica encontrada no experimento foi correlacionada com os dados obtidos com as análises *in silico*, utilizando-se os *primers* desenhados a partir da sequência desse *contig*. Foi verificado (Figura 6) que a expressão (relativo ao clone 22 irrigado) desse gene, nos clones de *C. canephora* testados, foi sempre maior na condição em que a planta estava sob o estresse hídrico, e também sempre maior nos clones considerados tolerantes à seca.

A construção da Base de Dados do Genoma Café, com sequências ESTs, foi de fundamental importância para a identificação de genes, com possível envolvimento na resposta ao estresse hídrico em cafeeiro.

Por meio das análises *in silico* apresentadas deste trabalho foi possível a identificação e seleção de vários genes candidatos à resposta ao estresse hídrico em cafeeiro, tais como genes envolvidos na resposta bioquímica na planta, incluindo acúmulo de osmólitos, detoxificação ou síntese de compostos de proteção e adaptações fisiológicas que podem elevar a eficiência do uso da água. A identificação e caracterização desses genes possibilitam a realização de estudos posteriores, tais como analisar a expressão diferencial entre plantas cultivadas em campo e casa de vegetação, ou a prospecção da variabilidade natural, utilizando-se diferentes materiais genéticos, que possibilitam assim a identificação e validação de polimorfismos, para o desenvolvimento de marcadores moleculares associados à tolerância à seca.



**FIGURA 5** – Dados de expressão relativa ao *cluster* mais expresso de cada biblioteca, para os 36 *clusters*, formados exclusivamente com ESTs das bibliotecas SH2 e SH3. A proporção de ocorrência de ESTs de cada biblioteca está ilustrada pelas diferentes cores.

**TABELA 6** – Resultados de BLASTX dos 36 *clusters*, formados exclusivamente com ESTs das bibliotecas SH2 e SH3.

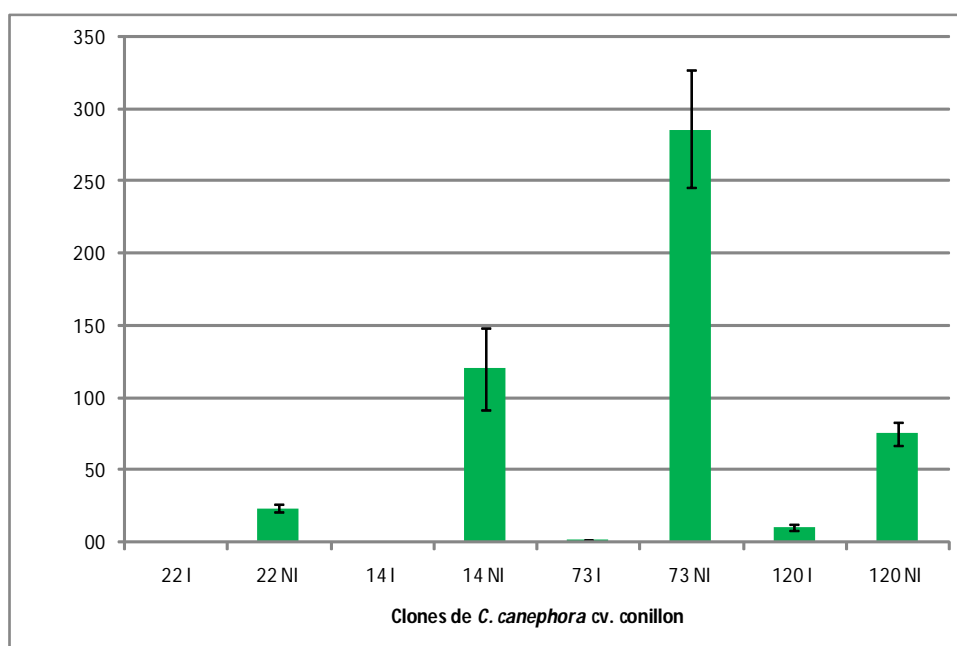
<i>Contig</i>	<b>Blast Hit   Função [Espécie em que foi identificada]</b>
<i>Contig18456</i>	gb AAV59263.1  At5g24670 [ <i>Arabidopsis thaliana</i> ]
<i>Contig18252</i>	gb AAT74541.1  dual-specificity phosphatase protein [ <i>Oryza sativa</i> ]
<i>Contig18249</i>	gb AAG50930.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig18165</i>	gb ABE78360.1  NAD-dependent epimerase/dehydratase [ <i>Medicago truncatula</i> ]
<i>Contig18142</i>	gb AAT70469.1  At2g46100 [ <i>Arabidopsis thaliana</i> ]
<i>Contig18082</i>	ref NP_565308.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig18080</i>	ref NP_197460.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig17989</i>	ref NP_192811.3  ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase [ <i>Arabidopsis thaliana</i> ]
<i>Contig17966</i>	dbj BAF17401.1  Os05g0401200 [ <i>Oryza sativa</i> (japonica cultivar-group)]
<i>Contig15296</i>	gb AAS67377.1  CONSTANS 1 [ <i>Solanum lycopersicum</i> ]
<i>Contig15292</i>	emb CAD10638.1  PBF68 protein [ <i>Nicotiana tabacum</i> ]
<i>Contig15115</i>	gb AAM19345.1  hypothetical protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig15078</i>	emb CAB85519.1  putative protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig15070</i>	gb AAM97542.1  catalase 3 [ <i>Capsicum annuum</i> ]
<i>Contig15003</i>	ref NP_193063.2  unknown protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig14994</i>	emb CAB80438.1  bZIP transcription factor-like protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig14951</i>	No hit
<i>Contig14915</i>	gb AAS76775.1  At1g34360 [ <i>Arabidopsis thaliana</i> ]
<i>Contig14907</i>	gb ABE85888.1  MtrDRAFT_AC148765g17v1 [ <i>Medicago truncatula</i> ]
<i>Contig14857</i>	dbj BAD87356.1  unknown protein [ <i>Oryza sativa</i> (japonica cultivar-group)]
<i>Contig14811</i>	dbj BAA98190.1  regulator of chromosome condensation like [ <i>Arabidopsis thaliana</i> ]
<i>Contig18009</i>	No hit
<i>Contig18007</i>	ref NP_179688.1  unknown protein [ <i>Arabidopsis thaliana</i> ]
<i>Contig17928</i>	gb AAO33924.1  phosphoenolpyruvate carboxylase kinase 1 [ <i>Solanum lycopersicum</i> ]
<i>Contig17887</i>	gb ABE92607.1  Agenet [ <i>Medicago truncatula</i> ]
<i>Contig17829</i>	ref NP_177561.1  glycerophosphodiester phosphodiesterase [ <i>Arabidopsis thaliana</i> ]
<i>Contig14831</i>	gb ABF00136.1  SNI1 [ <i>Solanum tuberosum</i> ]
<i>Contig14824</i>	No hit
<i>Contig14838</i>	ref NP_567928.1  carbonate dehydratase/ zinc ion binding [ <i>Arabidopsis thaliana</i> ]
<i>Contig18283</i>	ref NP_564334.1  unknown protein [ <i>Arabidopsis thaliana</i> ]

Continua...



TABELA 6 – Continuação...

Contig	Blast Hit   Função [Espécie em que foi identificada]
Contig15208	emb CAB79558.1  putative protein [ <i>Arabidopsis thaliana</i> ]
Contig15284	tpe CAI56334.1  TPA: isoflavone reductase-like protein 5 [ <i>Vitis vinifera</i> ]
Contig18467	emb CAA44820.1  heat shock protein 70 [ <i>Nicotiana tabacum</i> ]
Contig18015	gb AAD41409.1  cytosolic class II low molecular weight hsp [ <i>Prunus dulcis</i> ]
Contig18002	emb CAI43948.1  catalase [ <i>Populus deltoides</i> ]
Contig15055	No hit

FIGURA 6 – Análise de expressão do Contig 18100 em 4 diferentes clones de *C. canephora* irrigados (I) e não irrigados (NI).

#### 4 CONCLUSÕES

Após a *clusterização* e análises dos *contigs* formados, utilizando-se as ferramentas de bioinformática presentes na Base de Dados do Genoma Café, foi possível a identificação de *clusters* formados especificamente por ESTs, provenientes das bibliotecas de cDNA, construídas a partir de material vegetal submetido ao estresse hídrico.

A metodologia utilizada neste trabalho possibilitou a identificação de vários novos genes com expressão diferencial em resposta ao estresse hídrico, genes esses, nunca antes descritos (*No hits*), abrindo novas

possibilidades experimentais para a caracterização funcional destes genes. Como o realizado com o experimento de qPCR, em que foi possível verificar a expressão diferencial do *no hit* (*contig* 18100), que corroborou com a análise *in silico*.

#### 5 REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, G. S. C. et al. Análise da expressão de genes candidatos para a tolerância à seca, em genótipos de *Coffea arabica*, por PCR quantitativo. In: TALENTO ESTUDANTIL, 13., 2008, Brasília. *Anais...* Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2008. 1 CD-ROM.

CHO, S. H. et al. Proteome analysis of gametophores identified a metallothionein involved in various abiotic stress responses in *Physcomitrella patens*. **Plant Cell Reports**, Berlin, v. 25, n. 5, p. 475-488, Apr. 2006.

CRUZ, F. et al. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. **Molecular Breeding**, Dordrecht, v. 23, n. 4, p. 1380-1385, 2009.

CUSHMAN, J. C.; BOHNERT, H. J. Genomic approaches to plant stress tolerance. **Current Opinion Plant Biology**, Oxford, v. 3, n. 2, p. 117-124, Apr. 2000.

DAMATTA, F. et al. Effects of soil water deficit and nitrogen nutrition on water relations and photosynthesis of pot-grown *Coffea canephora* Pierre. **Trees Structure and Function**, Berlin, v. 16, n. 8, p. 555-558, Nov. 2002.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Recursos Genéticos e Biotecnologia. **Projeto Genoma Café**. Brasília, 2006. Disponível em: <<http://www.cenargen.embrapa.br/biotec/genomacafe>>. Acesso em: 10 fev. 2009.

EYMERY, F.; REY, P. Immunocytolocalization of CDSP32 and CDSP34, two chloroplastic drought-induced stress proteins in *Solanum tuberosum* plants. **Plant Physiology and Biochemistry**, New Delhi, v. 37, n. 4, p. 305-312, Apr. 1999.

FERRÃO, R. G.; FONSECA, A. F. A.; FERRÃO, M. A. G. Avaliação de clones elites de café conilon em condição de estresse hídrico no estado do Espírito Santo. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 1., 2000, Poços de Caldas. **Anais...** Brasília: Embrapa Café; Minasplan, 2000. p. 402-404.

GAO, Z.; LOESCHER, W. H. NADPH supply and mannitol biosynthesis: characterization, cloning, and regulation of the non-reversible glyceraldehyde-3-phosphate dehydrogenase in celery leaves. **Plant Physiology**, Bethesda, v. 124, n. 1, p. 321-330, Sept. 2000.

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

**Coffee Science, Lavras, v. 7, n. 1, p. 1-19, jan./abr. 2012**

GYÖRGYÉY, J. et al. *Medicago truncatula* nodule expressed sequence tags. **Molecular Plant Microbe Interact**, Saint Paul, v. 13, n. 1, p. 62-71, Jan. 2000.

HANDBERG, K.; STOUGAARD, J. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. **The Plant Journal**, Denmark, v. 2, n. 4, p. 487-496, July 1992.

HOEVEN, R. van der et al. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. **The Plant Cell**, Rockville, v. 14, n. 7, p. 1441-1456, July 2002.

HOUDE, M. et al. Wheat EST resources for functional genomics of abiotic stress. **BMC Genomics**, San Diego, v. 7, n. 149, p. 1136-1142, June 2006.

INGRAM, J.; BARTELS, D. The molecular basis of dehydration tolerance in plants. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 47, p. 377-403, June 1996.

INSTITUTE FOR GENOMIC RESEARCH. **Computational biology and functional genomics laboratory**. Washington, 2001. Disponível em: <<http://sgn.cornell.edu/>>. Acesso em: 10 fev. 2009.

INZE, D.; VANMONTAGU, M. Oxidative stress in plants. **Current Opinion in Biotechnology**, London, n. 6, p. 153-158, 1995.

IVANOVA, R. et al. Gene structure of *Drosophila* diaphorase-1 the diversity of the transcripts in female and male adult flies. **Plant Genetics and Breeding**, Berlin, v. 19, n. 2, p. 60-64, Mar. 2005.

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

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, 5., 2009, Vitória. **Anais...** Brasília: Embrapa Café, 2009. 1 CD-ROM.

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

MORIMOTO, R. I. Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. **Genes & Development**, Cold Spring Harbor, v. 12, n. 24, p. 3788-3796, Dec. 1998.

MYSTRUCTURED QUERY LANGUAGE. **MySQL**. New York, 1995. Disponível em: <<http://www.mysql.com/>>. Acesso em: 10 fev. 2009.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **Basic local alignment search tool**. Washington, 1988. Disponível em: <<http://www.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml>>. Acesso em: 10 fev. 2009.

NOGUEIRA, F. T. S. et al. cDNA arrays analysis and data mining of sugarcane response to low temperature. **Plant Physiology**, Rockville, v. 132, n. 4, p. 1811-1824, Aug. 2003.

OHLROGGE, J.; BENNINGT, C. Unraveling plant metabolism by EST analysis. **Current Opinion Plant Biology**, Oxford, v. 3, n. 3, p. 224-228, June 2000.

PERRET, E. et al. Improved differential screening approach to analyse transcriptional variations in organized cDNA libraries. **Gene**, Paris, v. 208, n. 2, p. 103-115, Feb. 1998.

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

RAMAKERS, C. et al. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. **Neuroscience Letters**, Limerich, v. 339, n. 1, p. 62-66, Mar. 2003.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular cloning, a laboratory manual**. 2<sup>nd</sup> ed. New York: Cold Spring Harbor Laboratory, 1989. 2222 p.

SCHÖFFL, F.; PRÄNDL, R.; REINDL, A. Regulation of the heat-shock response. **Plant Physiology**, Washington, v. 118, n. 4, p. 1135-1141, Dec. 1998.

SHEN, B. et al. Partial sequencing and mapping of clones from two maize cDNA libraries. **Plant Molecular Biology**, Netherlands, v. 26, n. 4, p. 1085-1101, Nov. 1994.

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

TELLES, G. P.; SILVA, F. R. da. Trimming and clustering sugarcane ESTs. **Genetics and Molecular Biology**, São Paulo, v. 24, n. 1/4, p. 17-23, Jan./Dec. 2001.

VETTORE, A. et al. The libraries that made SUCEST. **Genetics and Molecular Biology**, São Paulo, v. 24, n. 1/4, p. 101-107, Jan./Dec. 2001.

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. et al. Análise *in silico* de genes potencialmente envolvidos na resposta aos estresses abióticos, presentes na base de dados do Genoma Café. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 4., 2005, Londrina. **Anais...** Brasília: Embrapa Café, 2005. 1 CD-ROM.

\_\_\_\_\_. Identificação de genes candidatos para a tolerância à seca em cafeeiro, por meio de análises da expressão gênica com macroarranjos de cDNA e PCR quantitativo. In: TALENTO ESTUDANTIL, 13., 2008, Brasília. **Anais...** Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2008a. 1 CD-ROM.

\_\_\_\_\_. Identificação e caracterização de genes candidatos para a tolerância à seca em clones de café conilon (*Coffea canephora*). In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 5., 2009, Vitória. **Anais...** Brasília: Embrapa Café, 2009. 1 CD-ROM.

\_\_\_\_\_. Molecular characterization of candidate genes involved in drought tolerance in coffee plants. In: ASIC, 22., 2008, Campinas. **Abstracts...** Campinas: Unicamp, 2008b. 1 CD-ROM.

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

ZHANG, D. et al. Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum* L.). **Genetics**, Austin, v. 168, n. 2, p. 595-608, Oct. 2004.

**Coffee Science, Lavras, v. 7, n. 1, p. 1-19, jan./abr. 2012**