

CONSTRUÇÃO DO PERFIL DE EXPRESSÃO GÊNICA DA RESISTÊNCIA À SECA DO CAFÉ A PARTIR DE DADOS DE SEQUENCIAMENTO DE SEGUNDA GERAÇÃO¹

Ramon Vidal^{5,6}, Thierry Leroy², Fabien De Bellis², David Pot², Gustavo Costa Rodrigues³, Gonçalo Amarante Guimarães Pereira⁵, Alan Carvalho Andrade⁴, Pierre Marraccini^{2,4}

¹Trabalho financiado CIRAD (ATP 2007/01), INRA e Embrapa.

²CIRAD-UMR AGAP, Montpellier, FR,

³Embrapa Cerrados, Planaltina, DF, BR,

⁴Embrapa Recursos Genéticos e Biotecnologia (LGM-NTBio), Brasília, DF, BR,

⁵UNICAMP IB/LGE, Campinas, SP, BR,

⁶LNBio, CNPEM, Campinas, BR.

RESUMO: Para estudar os mecanismos moleculares envolvidos na resistência à seca em plantas de café, meristemas de ramos plagiotrópicos dos cultivares de *Coffea arabica*, IAPAR59 (I59 - planta com provável resistência à seca) e Rubi (R - planta com pouca resistência à seca) crescidas em condições normais (I) e de estresse hídrico (NI) foram coletados e usados para construção de bibliotecas de cDNA que foram sequenciadas usando a plataforma GS-FLX *Titanium* da Roche. Após tratamento dos dados brutos, foram obtidas 207.516, 109.005, 250.413 e 173.693 *reads* das amostras de I59-I, I59-NI, R-I e R-NI, respectivamente, totalizando mais de 255 Mb. Todas essas bibliotecas foram mapeadas no transcriptoma de *Coffea arabica* para identificação indireta da expressão de cada gene para cada condição. Do transcriptoma total formado por 55578 *contigs*, 48730 *contigs* tiveram *reads* mapeados e 7899 novos *contigs* foram formados. Com o banco de dados é possível identificar genes e grupos de genes expressos exclusivamente por uma espécie ou outra, em condição de estresse hídrico ou não. Esses são os primeiros dados de meristema de café submetidos a estresse hídrico, os resultados dessa comparação serão apresentados como dados preliminares de um *northern* eletrônico para identificação de genes diferencialmente expressos em ambos os cultivares em ambas as condições.

Palavras-chave: *Coffea arabica*, estresse hídrico, expressão gênica, 454, NGS

CONSTRUCTION OF A GENE EXPRESSION PROFILE OF DROUGHT RESISTENCE IN COFFEE USING SECOND GENERATION SEQUENCING¹

ABSTRACT: In order to study the molecular mechanisms involved in drought stress tolerance in coffee plants, meristems of plagiotropic branches of the cultivars of *Coffea arabica*, Iapar59 (I59 - probable plant with drought resistance) and Ruby (R - plant with low drought resistance) grown in normal conditions (I) and water stress (NI) were collected and used for construction of cDNA libraries that were sequenced using the GS FLX *Titanium* Roche platform. After processing the raw data were obtained 207,516, 109,005, 250,413 and 173,693 reads samples of I59-I I59-NI, RI and R-NI, respectively, totalizing more than 255Mb. All these libraries were mapped against the transcriptome of *Coffea arabica* to identify expression of each gene for each condition. From the whole transcriptome, consists of 55,578 contigs, 48,730 reads were mapped and 7899 new contigs were created. With this database it is possible to identify genes and groups of genes expressed exclusively by one species or another, on condition of water stress or not. These are the first data from meristem coffee subjected to drought; the results of this comparison will be presented as preliminary data of an electronic northern performed in order to identify differentially expressed genes in both cultivars under both conditions.

Key words: *Coffea arabica*, drought stress, gene expression, 454, NGS

INTRODUÇÃO

É bem conhecido que longos períodos de seca afeta o desenvolvimento de plantas de café, levando a planta a morte e aborto de frutos nos casos mais severos de seca, ou também afetando o florescimento ou desenvolvimento dos grãos no caso de estresse moderado (DaMatta et al, 2006). Em relação à diversidade genética do café, muitos trabalhos mostraram plantas de *C. canephora* var. Conilon com tolerância a seca (Fonseca et al, 2004; Ferrão et al, 2000) o que foi analisado posteriormente a nível fisiológico (DaMatta et al, 2003; Lima et al, 2002; Pinheiro et al, 2004; Pinheiro et al, 2005) e também utilizado para identificar genes candidatos a resposta a esse estresse (Vinecky et al, 2005). Apesar de estreita, a diversidade genética para tolerância a seca também existe em espécies de *C. arabica*. O desenvolvimento das tecnologias de sequenciamento de nova geração possibilitou a geração de grande quantidade de dados a um baixo custo e abriu um caminho para traçar perfis de expressão gênica diferencial comparando a frequência de reads obtidos após o sequenciamento. Para iniciar esse tipo de abordagem no café, foram construídas bibliotecas de RNA de tecidos a

partir de meristema dos cultivares IAPAR59 (I59, tolerante à seca) e Rubi (R, sensível à seca) cultivados em campo com (I) ou sem (NI) irrigação. Essas bibliotecas foram posteriormente sequenciadas e os resultados preliminares desse estudo são aqui apresentados.

MATERIAL E MÉTODOS

Os testes de campo foram realizados usando plantas de dois anos de idade dos cultivares IAPAR59 (I59) e Rubi MG1192. Sendo o primeiro considerado mais tolerante à seca do que o segundo (MAG Ferrão, comunicação pessoal), cultivado em condições de campo na Fazenda Experimental estação do centro da Embrapa Cerrado (Planaltina-DF, Brasil, 15°35'43"S - 47°43'52"O). Para ambos cultivares, as plantas foram cultivadas com (I) e sem (NI) irrigação durante a estação das secas. Para o sequenciamento, ápices dos ramos plagiotrópicos foram coletados entre 25 e 30 dias (entre 10:00 h e 12:00 h). As amostras foram armazenadas em etanol 75% RNase-free. Detalhes de experimento em campos e procedimentos de amostragem estão descritos em Rodrigues *et al*, 2011.

Extração de RNA

RNA total do ápice foi extraído conforme descrito previamente em Geromel *et al*, 2006.

Sequenciamento do cDNA

Sequências de cDNAs foram então sequenciadas realizando duas rodadas (uma biblioteca de cDNA por amostra x 2), utilizando GS-FLX Titanium (Beckman Coulter Genomics SA, Grenoble, França). Isso gerou cerca de um milhão de reads correspondentes a mais de 255 MB.

Análise de Bioinformática dos dados sequenciados

Todos os dados foram filtrados e trimados por critérios de qualidade, sequências de adaptadores utilizados no sequenciamento também foram retiradas após identificação pelo software SSAHA2 (Ning *et al*, 2001). O transcriptoma total de *C. arabica* foi utilizado como referência para o mapeamento dos reads de 454 (Mondego *et al*, 2011) utilizando o software MIRA (Chevreux *et al*, 2004). Scripts em PERL foram criados para contar o número de reads de cada biblioteca ancorados em cada contig a partir do arquivo ACE gerado na montagem. As frequências foram normalizadas para o tamanho total de cada biblioteca e as bibliotecas foram agrupadas para análise de expressão conforme Tabela 2 com o software IDEG6 (Romualdi *et al*, 2003). Os contigs foram anotados utilizando o software Blast2go (Conesa *et al*, 2006).

RESULTADOS E DISCUSSÃO

O transcriptoma de referência é formado por um total de 55578 clusters correspondendo a 54628 contigs e 950 singlets. Do total dos dados brutos de 454, cerca de 20% foram descartados devido à baixa qualidade. Os reads têm em média 236bp (depois da remoção do adaptador) e mais de 85% deles foram mapeados contra a referência (Tabela 1).

Library	T reads	F reads	A reads	T clusters	New Clusters	Exclusive Clusters	M reads	NC reads
TR	1166084	876503	-	55578	-	-	-	-
I59-NI	135304	109005	260	14906	710	273	76203	4024
I59-I	282213	207516	211	22660	1685	850	130805	7362
R-NI	230064	173693	233	20847	946	467	119697	4722
R-I	345751	250413	241	25122	1788	915	170490	7751
Total	993332	740627	-	-	-	-	497195	23859

Tabela 1 - Estatística de todos os reads mapeados contra o transcriptoma referência de *Coffea arabica* [REF]. Cultivares (R: Rubi e I59: IAPAR59) de *C. arabica* e condições de campo (I: irrigados e NI: não irrigados) estão indicados. Reads: totais (T reads), filtrados (F reads), tamanho médio dos reads em bp (A reads), mapeados no Transcriptoma de referência (M reads), que formaram novos contigs (NC reads). Também tem informações de todos os clusters formados por cada biblioteca levando em conta o transcriptoma de referência (T clusters) o número de novos clusters formados por cada biblioteca (New Clusters) e o número de clusters exclusivos de cada biblioteca (Exclusive Clusters). TR: Transcriptoma de referência.

Das combinações de bibliotecas utilizadas para encontrar genes diferencialmente expressos (Tabela 2) foram utilizados apenas os contigs que tinham quantidade de reads superior a 4 e desses selecionamos contigs únicos para cada combinação em ambas as bibliotecas e aqueles contigs com foldchange >2 ou <-2 e p-value <0,05.

Lib1	X	Lib2	Total	Lib1=0	Lib2=0	Lib1>Lib2	Lib1<Lib2
I59-I	X	I59-NI	2432	5	57	95	551
R-I	X	R-NI	4034	9	33	142	718
I59-I	X	R-I	4196	28	32	692	174
I59-NI	X	R-NI	2121	36	14	455	94
I59-I	X	R-NI	3006	18	49	404	464
I59-NI	X	R-I	3793	137	13	1024	31
I	X	NI	8951	4	58	332	1329
I59	X	R	8716	35	34	1172	263

Tabela 2 - Combinações de bibliotecas onde foram feitas comparações por meio de busca por genes diferencialmente expressos, em vermelho algumas situações de destaque onde podemos encontrar genes potencialmente ligados a resistência a seca.

Utilizamos as combinações “I59-I X I59-NI”, “R-I X R-NI”, “I59-I X R-I” e “I59-NI X R-NI” para buscar termos GO enriquecidos e encontramos entre os genes diferencialmente expressos comparando essas bibliotecas e o transcriptoma total. Por exemplo, genes envolvidos com resposta a estresse aparecem super-representados para o Rubi (cultivar susceptível à seca) em comparação com o IAPAR59 (Figura 1).

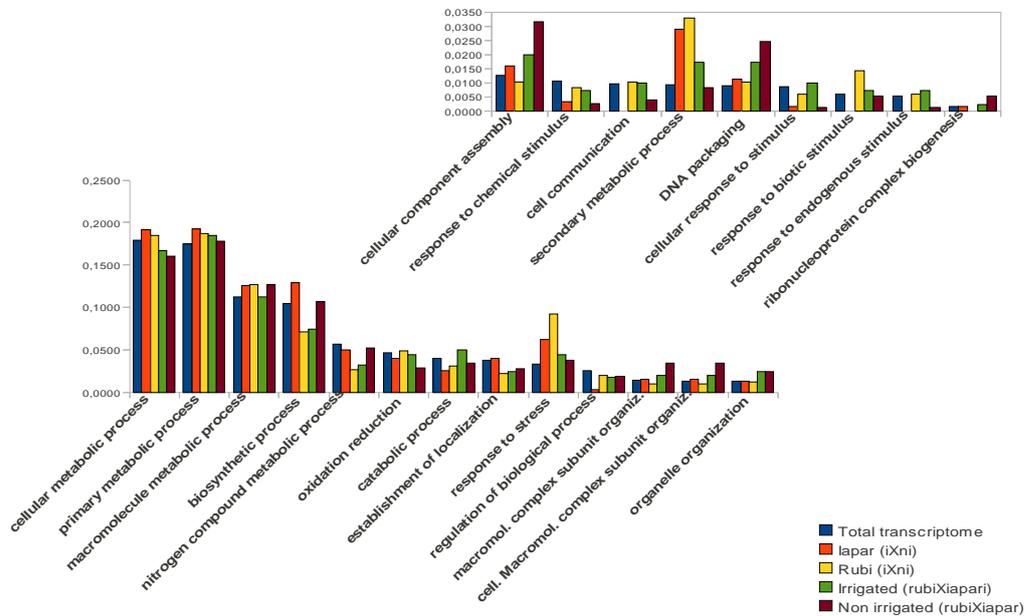


Figura 1 - Termos do Gene ontology (GO) enriquecidos para algumas combinações de bibliotecas em comparação com o transcriptoma total. O eixo Y indica o número de genes normalizados pelo número total de genes mapeados em cada biblioteca.

Isto pode estar relacionado com o fato de que o cultivar IAPAR59 apresenta Ψ_{pd} maior do que Rubi na colheita, demonstrando que o cultivar I59 tem maior eficiência no uso da água do que Rubi. É interessante notar também o número de genes com alta expressão em Rubi-NI e envolvido com resposta a estresse hídrico, como *RD22*, *PDIR10*, *MYB* e o codificando para cafeine syntase por exemplo. Estes genes não foram detectados I59 no momento de estresse. Por outro lado, muitos genes relacionados com a tolerância ao déficit hídrico foram detectados no cultivar I59, envolvidos no processo de biossíntese, como transportadores de açúcares, algumas proteínas relacionadas com a tolerância a estresses abióticos (ie osmotinas), proteínas de canal de água, proteínas LEA (desintoxicação e redução de danos celulares durante a desidratação) e proteínas de choque térmico (HSPs). No cultivar I59, genes que codificam enzimas envolvidas na síntese de ABA (isopentenyl diphosphate isomerase, geranylgeranyl redutase), o que provoca o fechamento dos estômatos, também foram altamente expressos.

CONCLUSÕES

O presente trabalho descreve os primeiros indícios de uma possível resistência à seca por uma cultivar de *Coffea arabica*. Em tempos de aquecimento global, com períodos de seca cada vez maiores, essa é uma preocupação eminente. Foram gerados quase 1 milhão de novas sequências de transcritos de *Coffea arabica*, das cultivares Iapar56 e Rubi, sendo descoberto cerca de 3 mil possíveis novos genes, alguns provavelmente exclusivos dessas cultivares. Os

dados gerais sugerem que a planta visivelmente mais suscetível à seca está superexpressando genes relacionados a resposta ao estresse e estresse hídrico, o que não é observado na planta resistente. Por outro lado, na planta resistente é observado uma superexpressão de genes de sinalização como hormônios e fatores de transcrição.

REFERENCIAS BIBLIOGRÁFICAS

- Chevreur, B., Pfisterer, T., Drescher, B., Driesel, A.J., Müller W.E.G., Wetter, T., Suhai, S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **2004**, 14, 1147-1159.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, 21, 3674-3676
- DaMatta, F.M., Ramalho, J.C. Impact of drought and temperature stress on coffee physiology and production: a review. *Braz. J. Plant Physiol.* **2006**, 18, 55-81.
- DaMatta, F.M., Chaves, A.R.M., Pinheiro, H.A., Ducatti, C., Loureiro, M.E. Drought tolerance of two field-grown clones of *Coffea canephora*. *Plant Sci.* **2003**, 164, 111-117.
- Ferrão, R.G., Fonseca, A.F.A., Silveira, J.S.M., Ferrão, M.A.G., Bragança, S.M. EMCAPA 8141 - Robustão Capixaba, variedade clonal de café conilon tolerante à seca, desenvolvida para o estado do Espírito Santo. *Rev. Ceres* **2000**, 273, 555-560.
- Fonseca, A.F.A., Ferrão, M.A.G., Ferrão, R.G., Verdin Filho, A.C., Volpi, P.S., Zucatei F. Conilon Vitória - Incaper 8142: improved *Coffea canephora* var. Kouillou clone cultivar for the State Espírito Santo. *Crop Breed. Appl. Biotech.* **2004**, 4, 1-3.
- Geromel, C., Ferreira, L.P., Cavalari, A.A., Pereira, L.F.P., Guerreiro, S.M.C., Vieira, L.G.E., Leroy, T., Pot, D., Mazzafera, P., Marraccini, P.. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *J. Exp. Bot.* **2006**, 57, 3243-3258.
- Lima, A.L.S., DaMatta, F.M., Pinheiro, H.A., Totola, M.R., Loureiro, M.E. Photochemical responses and oxidative stress in two clones of *Coffea canephora* under water stress. *Env. Exp. Bot.* **2002**, 47, 239-247.
- MONDEGO, J.M.C.; VIDAL, R.O.; CARAZZOLLE, M.F.; TOKUDA, E.K.; PARIZZI, L.P.; COSTA, G.G.L.; PEREIRA, L.F.P; ANDRADE, A.C.; COLOMBO, C.A.; VIEIRA, L.G.E.; PEREIRA, G.A.G.; AND BRAZILIAN COFFEE GENOME PROJECT CONSORTIUM. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. **BMC Plant Biology**, n. 11, p. 30, 2011.
- Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA. IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics*. 12(2):159-62, 2003
- Rodrigues *et al*, Anais do VII Simposio de pesquisa dos Cafés do Brasil. 2011
- Ning, Z., Cox, A.J., Mullikin, J.C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **2001**, 11, 1725-1729.
- Pinheiro, H.A., DaMatta, F.M., Chaves, A.R.M., Fontes, E.P.B., Loureiro, M.E. Drought tolerance in relation to protection against oxidative stress in clones of *Coffea canephora* subjected to long-term drought. *Plant Sci.* **2004**, 167, 1307-1314.
- Pinheiro, H.A., DaMatta, F.M., Chaves, A.R.M., Loureiro, M.E., Ducatti, C. Drought tolerance is associated with rooting depth and stomatal control of water use in clones of *Coffea canephora*. *Ann. Bot.* **2005**, 96, 1001-108.
- Vinecky F., de Brito, K.M., Da Silva, F.R., Andrade, A.C. Análise *in silico* de genes potencialmente envolvidos na resposta aos estresses abióticos, presentes na base de dados do projeto Genoma café. In *Proceedings of IV Simpósio de Pesquisa dos Cafés do Brasil*, Londrina-PR, Embrapa Café, CD-rom, 2005