

ANÁLISIS DE SECUENCIAS DE GENES DE *Coffea arabica* VAR. CATURRA

Gladis Estella Montoya-Ortiz*; Marco Aurelio Cristancho-Ardila**;
María del Pilar Moncada- Botero***

RESUMEN

MONTOYA O., G. E.; CRISTANCHO A., M. A.; MONCADA B., M del P. Análisis de secuencias de genes de *Coffea arabica* var. Caturra. *Cenicafé* 57(2): 79-87. 2006.

Actualmente en Cenicafé se están realizando estudios moleculares del café *Coffea arabica*, con el fin de identificar genes de importancia agronómica. En la presente investigación se desarrollaron genotecas de ADNc de tres tejidos diferentes (hoja, flor y fruto) de var. Caturra, con el fin de obtener información sobre los genes que se están expresando en estos tejidos. De 3.029 secuencias analizadas, se obtuvo un total de 1.824 transcritos únicos (unigenes), para las tres genotecas, y el 60% de estos transcritos mostraron similitud con secuencias reportadas para *C. canephora*. Los análisis de los clones identificados mostraron una amplia diversidad de genes involucrados en diversas funciones metabólicas. Las principales categorías incluyen proteínas involucradas en organización celular y biogénesis, metabolismo, síntesis de proteínas, transporte, transcripción, transducción de señales y respuesta a patógenos o plagas. Estos resultados concuerdan con algunos estudios realizados en tomate y *Arabidopsis* y pueden representar una tendencia general en plantas. Los transcritos involucrados en respuesta a patógenos, plagas y estrés abiótico, podrían ser usados en experimentos de laboratorio para confirmar su función y determinar su potencial de uso en programas de mejoramiento genético del café.

Palabras clave: *Coffea arabica*, genotecas de ADNc, ESTs, función proteica putativa.

ABSTRACT

Currently at Cenicafé molecular studies of coffee *Coffea arabica* are being carried out with the purpose of identifying genes of agronomic importance. In this work, cDNA libraries of three different tissues (leaf, flower and fruit) from *C. arabica* var. Caturra were developed with the objective of obtaining data about the genes expressed in these tissues. A total of 1,824 unique transcripts (unigenes) was obtained out of the 3,029 sequences analyzed for the three cDNA libraries, 60% of them showed similarity with sequences reported for *C. canephora*. The analysis of the identified clones showed a broad diversity of genes involved in different metabolic functions. The main categories include proteins involved in cell organization and biogenesis, metabolism, protein synthesis, transport, transcription, signal transduction and response to pathogens or pests. These results agree with some studies made in tomato and *Arabidopsis* and can represent a general tendency in plants. Transcripts involved in response to pathogens, plagues and abiotic stress could be used in laboratory tests to confirm their function and to determine their potential use in coffee genetic improvement programs.

Keywords: *Coffea arabica*, cDNA libraries, ESTs, putative proteic function.

* Bióloga MSc. Ciencias Microbiológicas.

** Investigador Científico II. Fitopatología, Centro Nacional de Investigaciones de Café, Cenicafé. Chinchiná, Caldas, Colombia.

*** Investigador Científico III. Mejoramiento Genético, Centro Nacional de Investigaciones de Café, Cenicafé. Chinchiná, Caldas, Colombia.

El café es uno de los principales productos agrícolas del mundo que genera ingresos a los países productores. Los estudios de esta planta han sido principalmente a nivel agronómico, sin embargo, el conocimiento a nivel bioquímico y molecular es aún limitado. Este conocimiento es importante para modificar o introducir características de interés agronómico en los programas de mejoramiento. El género *Coffea* ha sido estudiado a nivel agronómico y recientemente a nivel molecular principalmente en Brasil, Colombia y Francia.

Coffea spp. es una especie perenne, con $2X=22$ cromosomas para las especies diploides y $2X=44$ para la especie tetraploide *C. arabica*. Taxonómicamente, *Coffea* spp. pertenece a la familia Rubiaceae y es el único género de importancia económica en la familia. *C. arabica* es la especie más ampliamente cultivada de café en el mundo (75% de las plantaciones) y la única sembrada comercialmente en Colombia. Recientemente Lashermes *et al.* (12), han establecido el origen de *C. arabica* como un allopoliploide formado por la hibridación entre dos especies diploides cercanas (*C. eugenioides* y *C. canephora*). *C. arabica* se caracteriza por su baja diversidad genética, la cual es atribuida a su biología reproductiva y al proceso evolutivo de las especies. La baja variabilidad se refleja en la susceptibilidad de la especie a la mayoría de las enfermedades (4).

La generación de ESTs (Expressed Sequence Tags) a partir de genotecas de ADNc ha tomado gran importancia dentro de los estudios de genoma que actualmente se desarrollan en muchos organismos. Las secuencias ESTs se obtienen de la secuenciación parcial de clones de ADNc anónimos, presentes en genotecas de ADNc (1). Estos ESTs podrían representar idealmente todos los genes expresados en un tejido u órgano blanco, en un estado de desarrollo específico y/o bajo

condiciones medio ambientales específicas. Actualmente, más de 100 especies de plantas están representadas en la base de datos de ESTs (dbESTs) del GenBank (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), y resultan ser las plantas con mayor información las de las especies *Arabidopsis* sp, arroz, papa, maíz, tomate y soya.

El género *Coffea* cuenta con 47.999 registros de ESTs (Abril 2006), principalmente de secuencias de la especie *C. canephora*. Este tipo de proyectos permiten la caracterización de “sets” completos de transcritos de un organismo y son fuente de marcadores genéticos que pueden asociarse funcionalmente a características agronómicas de interés. Así, el uso de marcadores basados en ESTs puede conducir al mapeo genético de una secuencia blanco basada en la predicción de su función. En plantas, las genotecas de ADNc han permitido aislar y caracterizar clones que han contribuido a la construcción de mapas físicos y de ligamiento (7), al desarrollo de ESTs, así como también al estudio de los mecanismos de expresión de varias isoenzimas y varias familias de genes (19). En café, recientemente el grupo dirigido por Tanksley en la Universidad de Cornell (13), realizó un estudio con 47.000 secuencias de ADNc de *C. canephora* para establecer las posibles funciones de los transcritos y el perfil de expresión en diferentes tejidos, y comparar estos resultados con lo encontrado en *Arabidopsis* y Solanaceas (particularmente en tomate).

En Colombia, el proyecto del genoma del café tiene como metas fundamentales: conocer la localización de los genes de interés (resistencia a enfermedades y plagas, calidad de la bebida y rendimiento, entre otros), su secuencia (genómica estructural) y su función (genómica funcional). Dentro de esta última se enmarca el presente trabajo en el que se aislaron, caracterizaron y analizaron secuencias

de ADNc de tres tejidos distintos (hoja, flor y fruto) de *C. arabica* var. Caturra, con el fin de investigar de manera global, los perfiles de expresión génica en estos tejidos.

MATERIALES Y MÉTODOS

Material biológico. En la Estación Central Naranjal (Chinchiná, Caldas) se recolectaron hojas jóvenes, frutos de 22 semanas y botones florales de *C. arabica* variedad Caturra. Las muestras se tomaron directamente de plantas en el campo y los tejidos se conservaron inmediatamente en nitrógeno líquido para su transporte al laboratorio.

Aislamiento del ARN. De los tres tejidos se hizo la extracción del ARN total, usando el sistema de aislamiento para ARN total de Promega (RNAagents total RNA isolation system; Promega, Madison WI). Los ARNs mensajeros se obtuvieron del ARN total de cada tejido por separación magnética, empleando el sistema de aislamiento de Promega (PolyAtract mRNA isolation system; Promega, Madison, WI) y su integridad fue determinada por electroforesis en gel de agarosa.

RT-PCR. De 29 a 200ng de ARN mensajero fueron transcritos a ADN copia (ADNc) de cadena sencilla mediante la transcriptasa reversa “PowerScript” a 42°C por 1 hora. La primera cadena fue sintetizada con el cebador-adaptador oligo dT (CDCIII/3’ primer) y un adaptador que abarca la región 5’ del transcrito, para obtener transcritos de longitud completa. Esta cadena de ADNc sintetizada se amplificó por PCR con los cebadores CDCIII/3’ primer y 5’ PCR primer. La reacción se llevó a cabo con el siguiente perfil: 95°C por 15 segundos y 68°C por 6 minutos por 14, 21 y 24 ciclos para las muestras de hoja, flor y fruto, respectivamente.

Construcción de genotecas de ADNc. La síntesis de la doble cadena de ADN y la construcción de las genotecas se realizó siguiendo las especificaciones del fabricante del kit Creator Smart ADNc Library Construction Kit (Clontech, Palo Alto, CA, USA). Luego de la síntesis de la segunda cadena, el ADNc se digirió con la enzima *Sfi* I, se fraccionó por tamaños con columnas Chroma Spin 400 (Becton Dickinson) y se ligó al vector pDNR-LIB predigerido por los sitios *Sfi* IA y *Sfi* IB. Los productos obtenidos de las ligaciones se utilizaron para la transformación de bacterias *E. coli* DH10B.

Secuenciación de ADN. Las genotecas fueron procesadas por la compañía REXAGEN, Seattle WA. El proceso constó de plateo, selección de colonias, preparación del molde de ADN y secuenciación, los cuales fueron realizados en su totalidad por esta compañía. La secuenciación se realizó solo por el extremo 5’, con el cebador universal T7.

Análisis de secuencias. La depuración de las secuencias se realizó utilizando el programa CodonCode Aligner (versión 1.3.0). Con este programa se determinó la calidad de las secuencias (9) y aquellas secuencias con un valor de PHRED inferior a 30, fueron excluidas de los análisis. Posteriormente, se identificaron las secuencias correspondientes al vector, a los adaptadores y las secuencias contaminantes que, de igual manera, fueron omitidas para los análisis. Luego de esta depuración se procedió al ensamblaje de las secuencias con el programa CAP3 (10). El “set” de secuencias de unigenes se utilizó para realizar análisis de similitud con el programa Blastn (2) contra la base de datos de secuencias de unigenes de *C. canephora* (13), utilizando un “e-value” de 10⁻⁹. Adicionalmente, mediante el programa TargetIdentifier <http://fungalgenome.concordia.ca/tools/TargetIdentifier.html> (14), se procedió a

la determinación de la cantidad de transcritos de longitud completa. El programa busca características propias de secuencias completas (codones de inicio y de parada, ORFs) y anotaciones en marcos de lectura (ORFs), precedidos por el BlastX. Este BlastX se realiza contra la base de datos local de proteínas de *Arabidopsis* (29.161 secuencias).

Finalmente, para determinar la función putativa de los transcritos obtenidos, se realizó la traducción de la secuencia de nucleótidos a proteína mediante el programa ESTScan (versión 2.0) (11), el cual detecta regiones codificantes. Con el “set” de secuencias de aminoácidos, se procedió a realizar Blastp contra los unigenes encontrados en las secuencias de *C. canephora*, con un “e-value” igual al usado previamente.

La herramienta InterProScan (version 4.0) (20) se usó para la búsqueda de las posibles funciones putativas de las proteínas, por comparación con todas las bases de datos presentes en el InterProScan (Prosite, Prints. Pfam, ProDom etc.), arrojando como resultado un archivo con categorías del Gene Ontology (GO, <http://geneontology.org>).

RESULTADOS Y DISCUSIÓN

Generación de ESTs. Se secuenciaron 3.029 clones de las tres genotecas desde el

extremo 5' con el cebador universal T7. El número de secuencias obtenidas para cada una de las genotecas se encuentra resumido en la Tabla 1. Luego de la depuración de las secuencias correspondientes al vector, los adaptadores y las secuencias correspondientes a contaminación, se obtuvieron en total 2.208 secuencias (Tabla 1) listas para ser ensambladas con el programa CAP3. En este ensamblaje se obtuvieron 1.824 genes únicos, de los cuales 1.649 fueron singletons, 415 de hoja, 641 de flor, 593 de fruto y 175 contigs, con una longitud promedio de 400pb. Algunas secuencias presentaron poli A en su extremo 3', lo cual podría corresponder a transcritos completos dada la naturaleza de la síntesis del ADN copia. Para evaluar la cantidad de transcritos de longitud completa se corrió el programa Target Identifier, el cual determinó que de los 1.824 genes únicos el 19,68% corresponde a transcritos de longitud completa y el resto (80,32%) a secuencias de transcritos completos cortos, transcritos ambiguos y secuencias parciales (Tabla 2).

La redundancia estimada de las secuencias para cada genoteca se realizó con el programa wRedBlast (desarrollado en Cenicafé), y mostró que la genoteca de fruto contiene el mayor número de transcritos redundantes, con un valor alto del 32%, seguido de la genoteca de flor con un 21% y de la de hoja

Tabla 1. Descripción del número de secuencias aisladas y analizadas de los clones obtenidos en las Genotecas de ADNc tejido específicas de *C. arabica*.

	HOJA	FLOR	FRUTO	CONTIGS
Número de clones aislados	766	1.119	1.144	
Número de secuencias de buena calidad	557	845	806	
Porcentaje de redundancia	19	21	32	
Tamaño promedio de los insertos en pb ^a	431	388	350	466
Unigenes	415	641	593	175
Número de secuencias (proteínas ^b)	401	568	470	172

^a Pares de bases.

^b Secuencia de aminoácidos deducida a partir de la secuencia de nucleótidos.

Tabla 2. Predicción de las secuencias de café de longitud completa, obtenida con el programa Target Identifier.

DESCRIPCIÓN*	
Número total de secuencias	1.824
Secuencias de longitud completa	359
Secuencias de longitud completa cortas	76
Número de secuencias ambiguas	42
Número de secuencias parciales 5'	483
Número de secuencias parciales secuenciadas 3'	19

* Para las especificaciones, remitirse a los documentos descriptivos de las categorías del Target Identifier https://fungalgenome.concordia.ca/tools/docs/TargetIdentifier_faq.html

con tan sólo un 19% (Tabla 1). Por otra parte, no todas las secuencias de unigenes pudieron ser traducidas a proteínas, indicando transcritos con codones de parada a lo largo de la secuencia.

Similaridad de los transcritos de *C. arabica* con *C. canephora*. Una vez obtenidas las secuencias de unigenes, se realizó una búsqueda de similitud de las secuencias con las reportadas en *C. canephora* con el programa Blastn. Este análisis mostró que el 60% de las secuencias de *C. arabica* se comparten con las secuencias de la base de datos de *C. canephora*. En la Tabla 3 se describen aquellos transcritos que tuvieron similitud con los 20 unigenes más expresados en *C. canephora*.

Identificación de la función putativa. De los 1.824 transcritos iniciales, se obtuvieron 1.611 secuencias de aminoácidos y con ellas se realizó Blastp contra los unigenes de *C. canephora*. Este análisis mostró que el 47,9% de las secuencias de aminoácidos de *C. arabica* se encuentran reportadas para *C. canephora*. Dentro de las secuencias que mostraron similitud se encuentran transcritos identificados por Lin *et al.* (13) como específicos de desarrollo temprano, intermedio y tardío

del fruto; otros transcritos compartidos con *Arabidopsis* y otros transcritos únicos de café y solanáceas (Tabla 3).

Adicionalmente, las secuencias de aminoácidos fueron utilizadas para la identificación de las funciones putativas de las respectivas proteínas con el InterproScan. Los genes únicos que exhibieron similitud con proteínas reportadas en las diferentes bases de datos, se agruparon en diferentes categorías funcionales de acuerdo a la clasificación asignada por Gene Ontology¹ anotación (GO) para procesos biológicos. Se clasificaron dentro de las distintas categorías 371 secuencias (25%)(Figura 1), mientras que no fue posible asignarles una función específica a las 1.463 secuencias restantes, aunque muchas de ellas exhibieron dominios funcionales dentro de sus secuencias, pero sin posibilidad de identificarlas y clasificarlas en un proceso biológico específico dentro de la búsqueda en GO.

Las secuencias se ubicaron principalmente en las categorías de organización celular y biogénesis, metabolismo, síntesis de proteínas, transporte, transcripción, transducción de señales y respuesta a patógenos o plagas. Esta tendencia ya ha sido registrada para el

¹ GO. <http://www.geneontology.org>

Tabla 3. Relación de las proteínas de *C. arabica*, identificadas en la base de datos de *C. canephora*. E value > 10⁻¹⁵

DESCRIPCIÓN*	COMENTARIO*
IDENTIFICADOS A PARTIR DE NT¹	
Unigen 120912 (100% de similitud)	Proteína de almacenamiento 11S
Unigen 124988 (87% de similitud)	Función desconocida
Unigen 120685 (96% de similitud)	Quitinasa
Unigen 120121 (99% de similitud)	Taumatín, relacionada con patogénesis
Unigen 124574 (83% de similitud)	Proteína de resistencia a enfermedad
Unigen 126264 (84% de similitud)	Función desconocida (única de café)
IDENTIFICADOS A PARTIR DE AA²	
Factor de ribosilación-ADP	
Peroxidasa de secreción	
Metalotioneína	
Quitinasa	Implicada en resistencia a hongos
Sintetasa SAM	Específica de desarrollo temprano del fruto
Factor de transcripción WRKY4	Relacionado con estrés, infección por patógenos y senescencia. Específico de fruto
Sintetasa AdoMet	
Receptor plasmodesmal	
Subunidad pequeña de Rubisco	
Unigen 122071	Enzima clave en fotosíntesis y fijación de carbono
Unigen 125230	Proteína putativa S2 de almacenamiento
Unigen 124952	Compartida con <i>Arabidopsis</i>
Unigen 123451	Compartida con <i>Arabidopsis</i>
Unigen 120054	Compartida con <i>Arabidopsis</i>
Unigen 119449	Compartida con <i>Arabidopsis</i>
Unigen 122956	Retrotransposon. Único de café y solanáceas
Unigen 121998	Proteína de resistencia a enfermedad. Única de café y solanáceas
Unigen 123769	Proteína inducida ABA/WDS.
Unigen 120284	Proteína extension-like. Única de café y solanáceas.

* Tomado de Lin *et al.* (13). ¹. Nucleótidos. ². Aminoácidos

caso de *Arabidopsis* (3), tomate (17) y *C. canephora* (13), en los cuales se observó un comportamiento similar en los transcritos hallados. Estos resultados, junto con los

obtenidos en este trabajo, pueden indicar que este tipo de distribución o comportamiento es general para todas las plantas.

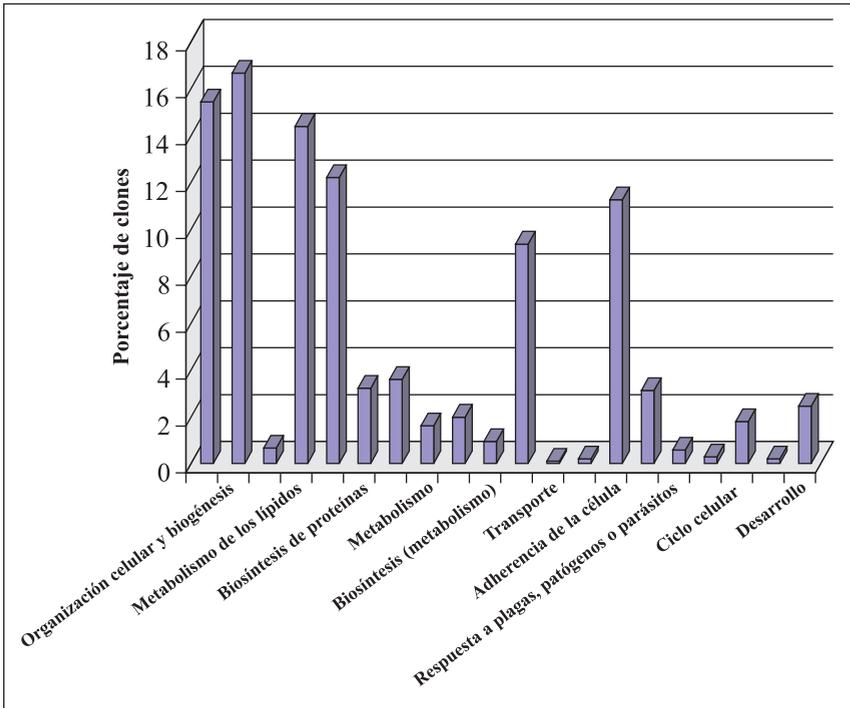


Figura 1. Distribución de las secuencias de unigenes de las tres genotecas con respecto a su función putativa de acuerdo a Gene Ontology.

Vale la pena resaltar los transcritos involucrados en respuesta a patógenos o plagas (3% en promedio); este valor es interesante por cuanto potencialmente se podría contar con transcritos que puedan estar involucrados en resistencia a ciertos patógenos, y en esa medida resulta importante el estudio futuro de estos transcritos dentro del programa de mejoramiento genético de la planta. Entre ellos se encuentran proteínas involucradas en la actividad de la quitinasa, la cual ha sido implicada en resistencia a hongos e insectos en varias especies de plantas, incluyendo café (5, 16), así como proteínas miembros de la familia SAR (Systemic Acquired Resistance), las cuales se han encontrado en ají y tabaco (Solanáceas) y parecen estar involucradas en el desarrollo de una resistencia sistémica adquirida a patógenos (18), luego de una respuesta hipersensible a una infección microbiana. SAR es caracterizada por una resistencia a largo plazo de un amplio rango de patógenos (18).

Se encontraron transcritos para proteínas implicadas en respuesta a estrés: Miembros de la familia de proteínas inducibles AB/WDS, que se producen por estrés hídrico (dehidrininas) (6, 15), peroxidasas que se producen por estrés oxidativo y catalasas que se sintetizan para neutralizar el peróxido de hidrógeno producido por diferentes organismos.

Finalmente, se detectó un transcrito para una proteína perteneciente a la familia MATE (“Multi Antimicrobial Extrusion”), la cual tiene homología con transportadores en bacterias. En *Arabidopsis*, la proteína *ALF5* (miembro de la familia MATE) se relaciona con resistencia a toxinas (8).

El estudio realizado permitió conocer parte del material genético del café, la posibilidad de realizar futuros análisis sobre la expresión génica tejido específica, identificación de genes con características agronómicas de interés y desarrollo de nuevos marcadores moleculares

(SNPs, CAPs y SSRs), que permitan construir el mapa genético del café.

Como conclusión se pueden resaltar los siguientes resultados:

- Los análisis mostraron que las genotecas contienen transcritos de redundancia moderada, los cuales son representativos e informativos.

- Los análisis de los ESTs identificados mostraron una amplia diversidad de genes involucrados en diversas funciones metabólicas. Las principales categorías incluyen proteínas involucradas en el metabolismo, en síntesis de proteínas y en rescate y defensa celular. Estas distribuciones concuerdan con los análisis realizados en *C. canephora* y algunos estudios realizados en tomate, donde se detectaron principalmente transcritos involucrados en procesos metabólicos, de transcripción y de síntesis de proteínas. Estas categorías también se encontraron en *Arabidopsis* y pueden representar una tendencia general en plantas.

- Se identificaron varios transcritos dentro de procesos de respuesta a patógenos y plagas, a estrés abiótico y a resistencia debida a toxinas. El análisis detallado de los mismos se hace llamativo debido a su potencial para incluirlos dentro de programas de mejoramiento genético de la planta de café.

- Las secuencias de ADNc analizadas en este estudio permitieron una caracterización inicial de varios genes codificantes dentro del genoma del café. Estos resultados pueden ser muy útiles en estudios de patrones de expresión génica, en el desarrollo de marcadores para la construcción de un mapa genético y la identificación de genes de defensa y resistencia en café.

AGRADECIMIENTOS

Los autores agradecen de manera especial a Luis Fernando Rivera y Carlos Eduardo Orozco de Cenicafé por sus aportes en las herramientas de Bioinformática para los análisis computacionales. Un reconocimiento muy especial al Ministerio de Agricultura y Desarrollo Rural de Colombia por la financiación de este proyecto.

LITERATURA CITADA

1. ADAMS, M.D.; KELLEY, J.M.; GOCAYNE, J.D.; DUBNICK, M.; POLYMERPOULOS, M.H.; XIAO, H.; MERRIL, C.R.; WU, A.; OLDE, B.; MORENO, R.F.; KERLAVAGE, A.A.R.; MCCOMBIE, W.R.; VENTER, J.C. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252(5013):1651-1656. 1991.
2. ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. *Journal of Molecular Biology* 215 (3): 403-410. 1990.
3. ASAMIZU, E.; NAKAMURA, Y.; SATO, S.; TABATA, S. A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Research* 7(3):175-180. 2000.
4. CALVALHO, A. Principles and practices of coffee plant breeding for productivity and quality factors: *Coffea arabica*. En : Clarke R.J., Macrae, R. (Eds.) Coffee. Vol 4: Agronomy. London, Elsevier Applied Science, 1988. p 129-165.
5. CHEN, Z.J. ; RIBEIRO, R. ; SILVA, M.C. Heat shock-induced susceptibility of green coffee leaves and berries to *Colletotrichum gloeosporioides* and its association to PR and hsp 70 gene expression. *Physiological and Molecular Plant Pathology* 63: 181-190. 2003.
6. CLOSE, T.J. ; KORTT, A.A.; CHANDLER, P.M. A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. *Plant Molecular Biology* 13: 95-108. 1989.

7. DAVIS, G.L.; MCMULLEN, M.; BAYSDORFER, C.; MUSKET, T.; GRANT, D.; STAEBELL, M.; XU, G.; POLACCO, M.; KOSTER, L.; MELIA H, S.; HOUCINHS, K.; CHAO, S.; COE, JR., E. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics* 152(3):1137-1172. 1999.
8. DIENER, A.C.; GAXIOLA, R.A.; FINK, G.R. Arabidopsis *ALF5*, a multidrug efflux transporter gene family member, confers resistance to toxins. *The Plant Cell* 13:1625-1638. 2001.
9. EWING, B; HILLIER, L; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* 8(3): 175- 185. 1998.
10. HUANG, X.; MADEM, A. CAP3: a DNA sequence assembly program. *Genome Research* 9(9): 868-877. 1999.
11. ISELI, C.; JONGENEEL, C.V.; BUCHER, P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proceedings of the International Conference Intell. Syst. Molecular Biology* 138-148. 1999.
12. LASHERMES, P.; COMBES, M.C.; ROBERT, J.; TROUSLOT, P.; D'HONTO, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. genome. *Molecular Genetics and Genomics*. 261 (2):259-266. 1999.
13. LIN, C.; MUELLER, L.A.; MCCARTHY, J.; CROUZILLAT, D.; PETIARD, V.; TANKSLEY, S. D. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theoretical and Applied Genetics* 112 (1): 114-130. 2005.
14. MIN, X.J.; BUTLER, G.; STORMS, R.; TSANG, A. TargetIdentifier: a web server for identifying full-length cDNAs from EST sequences. *Nucleic Acids Research* 33, Web Server Issue W669-W672. 2005.
15. PADMANABHAN, V.; DIAS, D.M.; NEWTON, R. J. Expression analysis of a gene family in loblolly pine (*Pinus taeda* L.) induced by water deficit stress. *Plant Molecular Biology* 35(6):801-807. 1997.
16. ROJAS H., R.; V. M.; LOYOLA V., V.M. Induction of a class III acidic chitinase in foliar explants of *Coffea arabica* L. during somatic embryogenesis and wounding. *Plant Science* 163(4):705-711. 2002.
17. VANDERHOEVEN, R.; RONNING, C.; GIOVANNONI, J.; MARTIN, G.; TANKSLEY, S. 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. *Plant Cell* 14(7):1441-1456. 2002.
18. VERBERNE, M.C.; VERPOORTE, R.; BOL, J.F.; MERCADO B., J.; LINTHORST, H.J. Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nature Biotechnology* 18(7):779-783. 2000.
19. YAMAMOTO, K.; SASAKI, T. Large-scale EST sequencing in rice. *Plant Molecular Biology* 35(1-2):135-144. 1997.
20. ZDOBNOV, E.M.; APWEILER, R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17 (9): 847-848. 2001.