

Generation of Expressed Sequence Tags (ESTs) for Gene Discovery and Marker Development in Cultivated Peanut

M. Luo, P. Dang, B. Z. Guo,* G. He, C. C. Holbrook, M. G. Bausher, and R. D. Lee

ABSTRACT

Expressed sequence tag (EST) libraries for cultivated peanut (*Arachis hypogaea* L.) were developed from two cDNA libraries constructed by means of mRNA prepared from leaves of peanut line C34-24 (resistant to leaf spots and *Tomato spotted wilt virus*) and immature pods of peanut line A13 (tolerant to drought stress and preharvest aflatoxin contamination). Randomly selected cDNA clones were partially sequenced to generate a total of 1825 ESTs, 769 from the C34-24 cDNA library and 1056 from the A13 cDNA library, in which 536 and 769 unique ESTs were identified, respectively. Results of BLASTx search showed that 52.8% of the ESTs from leaf tissue and 78.6% of the ESTs from the pod tissue have homology to genes of known function. Approximately 27.3 and 22.1% of ESTs matching homologous sequences in dbEST of GenBank on the basis of BLASTn algorithm have unknown functions. The ESTs were queried against MIPS functional catalog criteria and sorted according to putative function into 15 categories. A total of 1345 ESTs have been released to GenBank¹. Four hundred unigenes have been selected from these ESTs and arrayed on glass slides for gene expression analysis, and 44 EST-derived simple sequence repeat (SSR) markers have been characterized for cultivated peanut, in which over 20% of the SSRs produced polymorphic markers among 24 cultivated peanut genotypes. This is the first report of ESTs in cultivated peanut, and further characterization of resistance and stress genes may explain mechanisms functioning in these two peanut lines.

EXPRESSED SEQUENCE TAG libraries and databases have proven to be powerful tools for gene discovery, gene mapping, and for the analysis of quantitative traits. ESTs are generated by large-scale sequencing of randomly picked clones from cDNA libraries constructed from mRNA isolated at a particular development stage and/or tissue. Recently, the genomes of *Arabidopsis thaliana* (L.) Heyhn. (Martienssen, 2000) and rice (*Oryza sativa* L.) (Yamamoto and Sasaki, 1997; Yu et al., 2002) have been completely sequenced, and a large number of ESTs have become available in sequence databases. While the possible functions of many genes can be deduced by homologies to known genes in the databases, the functions of most of the emerging genes and regula-

¹ 1345 ESTs have been submitted to GenBank database under accession numbers: CD037499 to CD038843.

M. Luo and B. Z. Guo, USDA-ARS, Crop Protection and Management Research Unit, Tifton, GA 31793; M. Luo and R. D. Lee, Univ. of Georgia Coastal Plain Experiment Station, Tifton, GA 31793; P. Dang and M. G. Bausher, USDA-ARS, U.S. Horticultural Research Lab., Ft. Pierce, FL 34945; G. He, Center for Plant Biotechnology Research, Tuskegee Univ., Tuskegee, AL 36088; C. C. Holbrook, USDA-ARS, Crop Genetics and Breeding Research Unit, Tifton, GA 31793. Received 25 Sept. 2003. *Corresponding author (bguo@tifton.usda.gov).

Published in Crop Sci. 45:346–353 (2005).
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA

tory sequences are unknown (Somerville and Somerville, 1999).

Peanut is a member of the tribe Aeschynomere of family Fabaceae. Most of the species belonging to genus *Arachis* are self pollinated, but outcrossing can occur by bee pollination (Knauft et al., 1992; Stalker, 1997). Most *Arachis* species are diploid, but cultivated peanut is an allotetraploid ($2n = 4x = 40$) species native to South America (Krapovickas and Gregory, 1994). Although several *Arachis* species have been cultivated, *A. hypogaea* is the only one that has been both domesticated and is widely cultivated around the world in tropical, sub-subtropical, and warm temperate climates.

Peanut is an important crop internationally for both direct human food and oil production. Peanut yields are restricted in most areas of the world by diseases. Most peanut breeding programs are attempting to develop cultivars with improved disease resistance (Holbrook and Stalker, 2003). Both early and late leaf spot diseases (caused by *Cercospora arachidicola* Hori. and *Cercosporidium personatum* Berk. & Curt. Deighton, respectively) are among the worst foliar diseases of cultivated peanut. Leaf spot disease control in the USA has depended mainly on routine applications of chlorothalonil (tetrachloroisophthalonitrile), either on a calendar or advisory schedule (Branch and Culbreath, 1995). In peanut production areas in the southeastern USA, tomato spotted wilt disease caused by the *Tomato spotted wilt virus* (TSWV) has become more severe (Culbreath et al., 1999) and is a major yield-limiting factor. Control methods for TSWV are limited.

Industry and consumers have emphasized the food quality of peanut; however, peanuts are susceptible to *Aspergillus* infection, which can result in aflatoxin contamination (Holbrook et al., 1994). Peanuts also have proteins that result in allergic reactions (Li et al., 2000). Improvement of insect resistance, drought tolerance, oil quality, and flavor are also great challenges for breeding programs (Holbrook and Stalker, 2003).

Although an abundance of morphological variation within *A. hypogaea* is known, many agronomical traits are difficult to select by conventional selection techniques. This is in large part because most agronomically important traits in peanut are quantitatively inherited (Wynne and Coffelt, 1982), and significant genotype and environmental interactions exist. Molecular genetic research with peanut has mainly focused on molecular marker-assisted selection (Stalker and Mazingo, 2001) and genetic transformation (Ozias-Akins and Gill, 2001). Molecular markers are useful for crop improvement and studies of crop evolution in many species (Mohan et al., 1997). Unfortunately, little variation has been observed using molecular markers in *A. hypogaea* (Stalker and Mazingo, 2001), resulting in limited application. A reli-

able regeneration and transformation system in peanut has been established, and several genes for insect and virus resistance have been successfully introduced into peanut (Ozias-Akins and Gill, 2001). Although the application of transgenic technologies has enormous potential for enhancing trait improvement, the critical component needed for peanut cultivar development is identification of agronomically useful genes. Functional genomics research may aid in the development of polymorphic molecular markers (He et al., 2003) and provide useable genes for transgenic research.

With the development of the human genome project (HGP) and the establishment of genomics, new techniques to study gene expression in large scale and with high throughput were invented (Terry et al., 1999). The EST approach developed for use in the HGP, and it has been extensively applied in plant functional genome research (Somerville and Somerville, 1999). Other major crop genome projects were based on an EST strategy, including maize (*Zea mays* L., Gai et al., 2000), soybean [*Glycine max* (L.) Merr., Shoemaker et al., 2002], and wheat (*Triticum aestivum* L., Lazo et al., 2001). In the public database GenBank (NCBI), the total number of ESTs released in dbEST (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html; verified 7 Sep. 2004) was 15 808,211 (15 Feb. 2003); however, there were few peanut ESTs.

In the present research, two peanut genotypes were used to construct cDNA libraries and generate ESTs. One genotype (C34-24) was selected for its resistance to leaf spots and TSWV, and the other genotype (A13) was selected for its drought tolerance and reduced pre-harvest aflatoxin contamination. By analysis of the ESTs, we wanted to profile gene expression in leaf and immature-pod tissues, identify novel genes, and develop EST-derived SSR markers. This information can be used to identify genes for resistance and other traits, which may accelerate future breeding programs.

MATERIALS AND METHODS

cDNA Library Construction and Transformation

Two advanced peanut breeding lines of *A. hypogaea* used in this research were C34-24 and A13. C34-24 was selected for resistance to TSWV, and early and late leaf spot diseases. A13 was selected for drought tolerance and resistance to pre-harvest aflatoxin contamination. The plants used for cDNA construction were challenged by natural infestation. A total of 24 genotypes of cultivated peanuts were grown in the greenhouse and the genomic DNA was extracted for evaluation of EST-derived SSR polymorphism. Seeds from the two lines were field-planted (University of Georgia, Tifton) under standard cultural practices. Newly expanded leaves were bulked from over 30 plants of C34-24 at 100 d after planting (DAP). Immature pods at the R6 stage (Boote, 1982) were collected from over 20 plants of A13 at 100 DAP. All samples were immediately placed into liquid nitrogen and subsequently stored in a -80°C freezer.

Total RNA was isolated from leaves and pods with TRI-ZOL reagent (Invitrogen Corp., Carlsbad, CA) and mRNA was purified from total RNA with the PolyATtract mRNA isolation kit (Promega Corp., Madison, WI) following the

manufacturer's recommendations. The final concentration of mRNA was adjusted to $0.5 \mu\text{g}/\mu\text{L}$ with DEPC-treated water. Two cDNA libraries were constructed with mRNA ($5 \mu\text{g}$) from leaves or immature pods. Synthesis of cDNA and library construction were made following the protocol of the ZAP-cDNA Gigapack III Gold cloning kit (Stratagene, La Jolla, CA). Inserts were directionally cloned into Uni-ZAP XR vector by means of *XhoI* *EcoRI* site adapters. The λ library was packaged into λ phages with Gigapack III Gold. The library solution was diluted and used for transfection of the host bacteria XL1-blue and titer levels were calculated. The unamplified library was used to excise pBluescript phagemids from the Uni-ZAP XR vector according to the in vivo mass excision protocol provided by the manufacturer, and the phagemids were used to transform the host bacteria SOLR. The white clones grown on LB screening culture medium (Amp/IPTG/X-gal) were recovered by random colony selection.

Plasmid DNA Purification and DNA Sequencing

Plasmid DNA was isolated from randomly selected colonies with 96 TurboFilter Miniprep kits (Qiagen, Valencia, CA) and a Qiagen BioRobot 9600. Ninety-six individual plasmids from random selected clones were used as template for PCR amplification of the cloned cDNA by with T3 and T7 universal primers. The resulting amplicons were analyzed by agarose gel electrophoresis to identify insert size. The PCR products were concentrated and washed by ethanol precipitation and pellets were resuspended in $15 \mu\text{L}$ of formamide before sequencing. *Taq* polymerase sequencing reactions were performed with the BigDye Terminator Cycle Sequencing Kit (PerkinElmer, Foster City, CA) from the 5' end of the cDNA clone with the T3 primer. Automated sequencing was performed on an ABI Prism 3700 Sequencer (PE Applied Biosystems, Foster City, CA) at the USDA-ARS Genomics Laboratory, Fort Pierce, FL.

EST Processing and Sequence Analysis

Sequences were edited by the software SEQUENCHER V4.1.4 (Gene Codes, Ann Arbor, MI) to remove vector sequence. Vector trimming was also edited manually when abnormal sequences were observed. Sequences <150 nucleotides were removed. The remaining ESTs were submitted for comparison against the GenBank non-redundant protein database by BLASTX algorithm (Altschul et al., 1990). A match was considered significant when the score was higher than 100 (optimized similarity score) with *E*-value scores $\leq 10^{-10}$. Novel ESTs were identified by comparison with sequences in non-redundant nucleotide and EST databases of GenBank using BLASTn algorithm. The criteria to define nucleotide identity referenced to over 95% identity for more than 100 bp were included in the same group. In addition, the sequences of each contig were aligned by the fragment assembly program of SEQUENCHER, and consensus sequences were generated with 90% identity over a minimum of 50 nucleotides.

EST Functional Analysis and Simple Sequence Repeat Characterization and Polymorphism Identification

Putative functions of the ESTs were classified according to the Munich Information Center for Protein Sequences (MIPS) functional classification system applied to *Arabidopsis* (Mewes et al., 2002; Schoof et al., 2002). A total of 1345 ESTs from the two cDNA libraries were submitted to the GenBank database under accession numbers: CD037499 to CD038843, and these sequences were searched for simple SSRs by BLASTn soft-

Table 1. Primer sequences of selected gene from the leaf and immature-pod cDNA libraries for RT-PCR analysis.

Gene probe	Primer (5' to 3')	PCR product (bp)
1 H3.3	F: A G A T T T C A G A G C C A C G C A G T C R: G T C C T T G G G C A T A A T G G T G A C	123
2 Allergen Arah3/Arah4 (seed)	F: A A G A G C C T G C A C A A C A A G G A C R: G A G A T C A C C C T C A T C G A A A C G	128
3 10-kDa protein precursor (leaf)	F: T G C A A A G T G A G G C A G C A A C R: A C C A A C A G C G G A A A T C G T C	141
4 Leucine-rich receptor-like protein kinase (leaf)	F: C C A T G A G A A G G G T G G T G A A G R: A A G C A A C A C T T C C G T G A T C C	126
5 Bax inhibitor-1 like (seed)	F: T C T C T C C C G T C G T T C A G A A T C R: T G C A T C C A G C T G A C G T A A G A A	132
6 Cell-autonomous heat shock cognate protein 7 (leaf)	F: C A G T G G C T A G A C A G C A A C C A R: A G C T C C A C C A A C A T C A G G A C	132
7 Nonspecific lipid-transfer protein (leaf)	F: A G G A A G C G G A T A C C A C A C A G R: T C G C C T A A G T G G A G G T C C T A	131
8 Drought inducible 22-kD protein (leaf)	F: C A T G C T C A C A G G C A C A A G A T A R: G A G G T G G T G C T T C T T C T T G C	146
9 Drought-induced protein RPR-10 (seed)	F: G C C C T G G A A C T G T C A A G A A G R: C T C T T G G A A C C C T G T T C C T C	134
10 Defensin protein (seed)	F: T A A T G G C A T C G C T C T C T T C C R: A C A C T C C T T T G A A G C G A T G G	139

ware. The di- and trinucleotide and some tetranucleotide SSR motifs were used in the sequence analysis. SSR motifs, which repeat more than seven times in dinucleotide, five in trinucleotide, and four in tetranucleotide were counted. Repeats of dimeric, trimeric, and tetrameric motifs were also searched in the EST sequences. Primer pairs were designed for 44 SSRs on the basis of the number of repeats and the sequences of the flanking region by the Primer3 software (Rozen and Skaletsky, 2000) (code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html; verified 5 August 2004). Twenty-four cultivated peanut genotypes were screened for polymorphisms. Four hundred unigenes were arrayed on glass slides for gene expression analysis (Luo et al., 2003b).

RT-PCR Analysis of Genes with High Redundancy of Expression

Peanut lines of C34-24 and A13 were planted in the greenhouse. Three-month-old plants of C34-24 were inoculated with fungal spores of late leaf spot disease (2×10^5 spores/mL) (Zhang et al., 2001). A13 was inoculated with *A. flavus* under drought stressed conditions (Anderson et al., 1996). Bulk samples of leaves and immature pods were collected for total RNA extraction.

Total RNA was treated with RNase-free DNase (Qiagen) to remove the genomic DNA contamination and the first strand of cDNA was synthesized from 2 µg total RNA with SuperScript II RT (Invitrogen) and gene specific primers (Table 1) selected from the two cDNA library EST sequences with higher redundancy, according to the manufacturer's instructions. PCR products were amplified in 20 µL of reaction mixture, containing 1.5 µL of the first strand cDNA, 0.5 µM of each primer (Table 1), 1× PCR buffer, 0.4 mM dNTP, and 1 U of *Taq* DNA polymerase (Promega), according to the following protocol: 94°C for 2 min; 35 cycles at a temperature of 94°C for 15 s, 59°C for 30 s, 72°C for 30 s; final-elongation occurred at 72°C for 7 min. All the PCR reactions were repeated three times.

RESULTS AND DISCUSSION

Quality of cDNA Libraries and ESTs

Two cDNA libraries were constructed, one from bulked leaf samples of C34-24 and one from bulked immature pod samples of A13. Two thousand clones were randomly picked from each of the two cDNA libraries. The

plasmids isolated from 96 individual clones were used as templates for PCR amplification using T3 and T7 primers to estimate the insert size. In the leaf cDNA library, the insert size distribution ranged from 200 to 1500 bp, with an average of 550 bp. In the immature pod cDNA library, the insert size distribution ranged from 400 to 1500 bp, with an average of 650 bp.

Sequence analysis showed 769 high quality ESTs from the leaf library and 1056 from the immature-pod library. Three hundred ninety-one of the 769 leaf ESTs and 378 of the 1056 immature-pod ESTs were assembled into contigs, resulting in 536 and 800 unique ESTs in the two libraries, respectively.

Identification of ESTs' Putative Function

To identify the putative function of these ESTs, their sequences were compared with the sequences in the UniGene (Schuler, 1997; Wheeler et al., 2003) database of GenBank by a BLASTx algorithm. A nucleotide homology search indicated that 52.8% of the clones from the leaf library and 78.6% of the clones from the immature pod library matched known function genes. A further search in dbEST database of GenBank based on BLASTn algorithm indicated that 27.3% of 363 clones from the leaf library and 22.1% of 226 clones from the immature pod library with unknown function matched sequences in dbEST database.

On the basis of the MIPS Functional Catalogue criteria (Mewes et al., 2002; Schoof et al., 2002) and the putative functions, the EST sequences in the two cDNA libraries were further characterized by functional category sorting into 15 groups (Fig. 1). Each putative transcript was assigned to a category in the MIPS Functional Catalogue (Mewes et al., 2002). The metabolism-related genes accounted for 6.6% of the cDNA clones in the leaf library and 6.8% in the immature-pod library (Fig. 1). In the protein synthesis category, cDNA clones encoding various ribosomal proteins and other formations from simpler components of a protein, rather than of proteins in general, were common in both libraries (Fig. 1). The cDNA clones related to cellular transport and transport mechanisms were low in both libraries (0.1%).

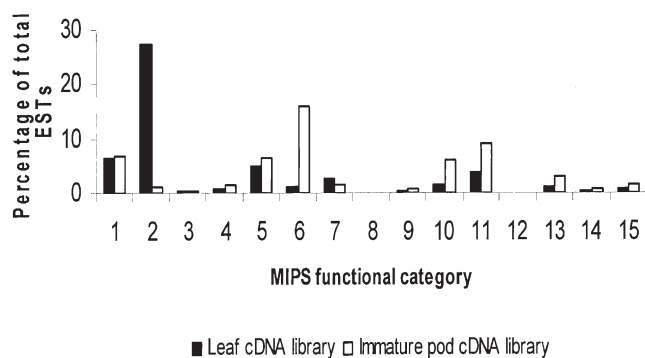


Fig. 1. Distribution of ESTs from leaf and immature cDNA libraries among functional categories. MIPS functional categories: 1-Metabolism; 2-Energy; 3-Cell growth, division and DNA synthesis; 4-Transcription; 5-Protein synthesis; 6-Protein destination; 7-Transport facilitation; 8-Cellular transport and transport mechanisms; 9-Cellular biogenesis; 10-Cellular communication/signal transduction mechanism; 11-Cell rescue, defense and virulence; 12-Ionic homeostasis; 13-Cellular organization; 14-Development; 15-Unclassified protein.

Differences were also seen in these two libraries in gene distribution among functional categories (Fig. 1). Energy-related genes were very common (27.3%) in the leaf cDNA library. This reflects the importance of photosynthesis in leaf tissue. In the immature-pod library, the most common types of genes were "Protein Destination." This finding was consistent with the function of the seed as a storage organ for both oil and protein in peanut. No EST function was categorized as "Ionic Homeostasis" in the leaf library, but 0.1% of the clones in the immature-pod library were grouped into this category.

ESTs related to signal transduction and cell defense from the two different peanut lines were expected to be seen. These genes were relatively abundant in both type and number (Fig. 1). In addition to these functional categories, there were also a proportion of ESTs which match their putative proteins. Some ESTs with unknown function matched homologous sequences in GenBank, and they were compared with sequences in dbEST (Boguski et al., 1993) of GenBank based on the

BLASTn algorithm. About 27.3% of 363 leaf ESTs and 22.1% of 226 immature-pod ESTs had homologous sequences with known genes. Some ESTs in both cDNA libraries matched with sequences from stress induced cDNA libraries, including stress factors of pathogens, ABA, drought, salt, cold, acid treatment, etiolated, irradiated, and phosphate-free culture.

Genes with Higher Redundancy of Expression

Redundancy of expression is a reflection of the importance of certain genes in different developmental stages and under different environmental conditions. Because of the small number of ESTs generated from these two libraries, ESTs with a redundancy of four and above were selected in this analysis (Table 2). In the leaf library, genes related with photosynthesis were ubiquitous among the redundant genes, and 13.9% of total ESTs were coded for Rubisco (ribulose biphosphate carboxylase) (Table 2). Rubisco is very abundant in chloroplasts, comprising more than 16% of the total protein in this organelle (Miziorko and Lorimer, 1982).

Several other genes reported to have function in adversity tolerance were found in the peanut leaf cDNA library. These include a nonspecific lipid-transfer protein, a 10-kDa protein precursor, caffeic acid *O*-methyltransferase, and a drought inducible 22-kDa protein. A nonspecific lipid-transfer protein has been reported to function in disease-resistance, and drought and cold-tolerance in plants (Park et al., 2002; Trevino and O'Connell, 1998; Pearce et al., 1998). The 10-kDa protein precursor has antifungal activity (Ye and Ng, 2002). Richard et al. (1996) reported that caffeic acid *O*-methyltransferase plays a role in the secondary metabolism-phenylpropanoid pathway and conditions disease resistance.

Several ESTs encoding candidate allergens were found in the immature-pod cDNA library (Table 2). These include the storage proteins glycinin and conglutin, each accounting for over 5% of the total ESTs (Table 2). These two storage proteins, along with the Ara h 1, are known peanut allergens (Mills et al., 2001; Xiang et al.,

Table 2. Redundant genes in the leaf and immature pod cDNA libraries of peanut.

Homologous group	Leaf		Immature-pod	
	Putative genes	No.	Putative genes	No.
1	ribulose biphosphate carboxylase	107	glycinin	54
2	chlorophyll a/b-binding protein	13	conglutin	53
3	ferredoxin i	13	mannose/glucose-binding lectin precursor	32
4	fructose-biphosphate aldolase 1	12	metallothionein	27
5	oxygen-evolving enhancer protein	11	allergen Ara h 1	17
6	nonspecific lipid-transfer protein	11	oleosin	16
7	plastocyanin	7	drought-induced protein RPR-10	9
8	glyceraldehyde 3-phosphate dehydrogenase	7	auxin-repressed protein	9
9	no similarity	6	no similarity	8
10	10-kDa protein precursor	6	catalase	7
11	photosystem I subunit IV	6	60 ribosome protein L10	6
12	caffeic acid <i>O</i> -methyltransferase	5	no similarity	6
13	ribulose-biphosphate carboxylase activase	5	defensin protein precursor	5
14	no similarity	5	cyclophilin	4
15	function unknown	5	fructose-biphosphate aldolase	4
16	photosystem II 10-kDa polypeptide	4	ubiquitin-conjugating enzyme	4
17	drought inducible protein	4	proteinase inhibitor	4
18	endopeptidase Clp	4	extensin-like protein	4
19	aldo/keto reductase	4	F-box protein family	4
20	no similarity	4	HSP80	4
21			no similarity	4

Table 3. Putative adversity resistance genes in the leaf and immature-pod cDNA libraries of peanut.

Class of function	Leaf	Immature-pod
Secondary metabolism	caffeic acid <i>O</i> -methyltransferase, cinnamoyl-CoA reductase	1-aminocyclopropane-1-carboxylate oxidase like protein, phenylalanine ammonia-lyase 3, cinnamoyl-CoA reductase
Reactive oxygen scavengers	glutathione <i>S</i> -transferase (GST 8, GST 9), superoxide dismutase [Cu-Zn],	glutathione <i>S</i> -transferase GST 9, Catalase, Catalase isozyme 1, superoxide dismutase (Cu-Zn), phytochelatin synthetase-like protein, ascorbate peroxidase
Stress proteins	salt tolerance-like protein, nitrate-induced protein, drought inducible protein, cytochrome p450, proline-rich protein, low temperature and salt responsive protein LTI6B, ClC protein, ultraviolet-B-repressible protein, aluminium induced protein, auxin-induced protein	drought-induced protein RPR-10, dehydration stress-induced protein, cytochrome P450, wound-induced protein, hydroxyproline-rich glycoprotein (sbHRGP3), NOI protein, cold-regulated LTCOR12, osmotin-like protein, late embryogenesis like abundant protein, selenium binding protein, fasciclin-like arabinogalactan-protein, fatty acid hydroxylase, auxin-regulated
Heat shock proteins	heat-shock protein 80	HSP70, HSP80, HSP90, 20.5-kDa HSP
Ion/proton transporters	Ca ²⁺ /H ⁺ -exchanging protein, H ⁺ -transporting ATPase, ankyrin	aquaporin-like transmembrane channel protein, dnaK-type molecular chaperone BiP precursor, vacuolar H ⁺ -ATPase catalytic subunit
Membrane fluidity	lipid transfer protein, nonspecific lipid-transfer protein, fatty acid desaturase	lipid transfer protein, nonspecific lipid-transfer protein
Signaling components	kinase C inhibitor, calcium binding protein, ethylene-responsive small GTP-binding protein, leucine-rich receptor-like protein kinase, Mannose/glucose-binding lectin precursor	rac-type small GTP-binding protein, casein kinase, Calmodulin, mannose/glucose-binding lectin precursor, galactose-binding lectin precursor, lectin-like protein, protein receptor-like kinase homolog, leucine-rich repeat transmembrane protein kinase, calmodulin-related protein
Control of transcription	14-3-3-like protein A, zinc finger protein	myb family transcription factor, Myb-related transcription factor-1, bZIP transcription factor 6
Defense response	10-kDa protein precursor, peroxidase, leucine-rich repeat protein LRP, beta-glucosidase, metallothionein, AIG2-like protein, glycosyl hydrolase family 19, senescence-associated protein, phytoene dehydrogenase, allyl alcohol dehydrogenase	trypsin inhibitor pathogenesis-related protein (PR-1, PR10), defensin protein precursor, proteinase inhibitor se60-like protein, Kunitz-type trypsin inhibitor, Bowman-birk type proteinase inhibitor A-II, protease inhibitor, Metallothionein, glycosyl hydrolase family (3,19), peroxidase, alcohol dehydrogenase, 1-Cys peroxiredoxin, systemin receptor SR160, blue copper protein, lipoxigenase 1, auxin-repressed protein

2002). An oleosin is also a putative allergen (Pons et al., 2002). Mannose/glucose-binding lectin precursor, present in the immature-pod cDNA library, is widely distributed in higher plants and is believed to play a role in recognition and binding of foreign high-mannose type glycans from microorganisms or plant predators (Barre et al., 2001). The defense-related genes were also present with high frequency, such as metallothionein, catalase, proteinase inhibitor se60-like protein, and F-box protein (Table 2). Ubiquitin-conjugating enzyme has been known to play a role in cell protection metabolism (Hellmann and Estelle, 2002) and was also present in this library.

Drought-induced genes were observed with redundancy in both cDNA libraries. This observation may be related to the level of drought stress in the field when the tissue samples were collected for cDNA library construction. The types of drought-induced genes were different in leaf and immature-pod cDNA libraries (Table 2). In addition, the functions of several redundant ESTs from these two cDNA libraries could not be identified from the public database, and most of these ESTs with unknown function do not have homologous sequences in GenBank. These ESTs may represent unique genes in peanut.

Putative Adversity Tolerance Genes

To identify genes that control tolerance to drought and disease resistance traits, the ESTs from these two cDNA libraries were characterized according to their putative functions related to previously reported adver-

sity resistance genes (Table 3). These genes were sorted into different categories and distributed among many defense-related pathways (Table 3). The disease resistance related genes are abundant, including leucine-rich repeat transmembrane protein kinase in pathogen recognition, glycosyl hydrolase family, pathogenesis-related protein (PR-1, PR10), glutathione *S*-transferase (GST 8, GST 9), β -glucosidase, and defensin protein precursor. Abiotic stress-related genes are also present in these ESTs (Table 3). Drought induced genes are in high proportion, such as dehydration stress-induced protein, osmotin-like protein, and hydroxyproline-rich glycoprotein (sbHRGP3). Other stress induced genes were also found in these ESTs, such as genes induced by low temperature and salt, auxin-induced, wound-induced, ultraviolet-B-repressible, and aluminum-induced proteins. Cytochrome P450 and heat-shock proteins have been reported as adversity resistance (Chapple, 1998; Sun et al., 2002). Signal transduction as influenced by abiotic and biotic factors has been a focus in searching the adversity resistance mechanisms in plant (Nurnberger and Scheel, 2001; Zhu, 2002). Several ESTs found in these two peanut cDNA libraries had homology to signal transduction related compounds (Table 3), but their roles in the signal transduction and defense-related pathways in peanut are not clearly understood.

EST-Derived SSR Markers and Gene Discovery or Expression

During the last decade, microsatellites or SSRs have proven to be useful markers in plant genetic research

Table 4. SSR (simple sequence repeat) motif and frequency in peanut ESTs.

Leaf		Immature pod	
SSR motif	Frequency	SSR motif	Frequency
(AT)7,11	6	(AT)7,8,11,17	6
(CT)8,9,11,17	7	(GA)8,10,19	5
(GAA) 5	1	(AAG)6,9	2
(CAA)5	4	(ACA)6,9	3
(AAT)5	5	(AAT)5,6,7	10
(TAT)6	2	(TAT)5,6,7,10,26	6
(GT)5	1	(CAC)5,6	4
(TTC) 5	1	(TTC)6	2
(ATC) 5	2	(ATC)5	4
		(AGC)5,6	2
		(CCT)6	1

and have been used for marker-assisted breeding purposes. Unfortunately, development of SSR markers is expensive, labor intensive, and time consuming if SSRs are being developed from genomic libraries (He et al., 2003). The cost of mining EST libraries is far lower than other traditional methods, and SSR development from ESTs has been successful in EST data mining (Cordeiro et al., 2001; Kantety et al., 2002). As determined by currently available molecular techniques, domesticated peanuts have a narrow genetic diversity (Stalker and Mazingo, 2001). Hopkins et al. (1999) tested the effectiveness of SSR markers in detecting molecular variation in cultivated peanut and they reported that the development of SSRs should enable the measurement of molecular polymorphism in the cultivated peanuts. After screening the ESTs from the two cDNA libraries by BLASTn, di- and tri- nucleotide and several tetranucleotide SSR motifs were found and used in the sequence analysis (Table 4). SSR motifs, which repeat more than seven times in dinucleotide, five in trinucleotide, and four in tetranucleotide were counted. In the expressed sequences there were only three types of dimeric repeat motifs [(GA)_n, (CT)_n, and (AT)_n] (Table 4). Trimeric repeat motifs are relatively higher than dimeric repeats. Tetrameric and hexameric repeat motifs were found in our ESTs. SSR markers have been detected in genomic DNA for dinucleotide repeats [(GT)₁₀ and (CT)₁₀] and the discrimination power of using SSRs as markers was high in domesticated peanuts (Hopkins et al., 1999; He et al., 2003). EST-derived SSR motifs were identified in these peanut ESTs, and 44 SSRs from the released 1350 ESTs (GenBank accession number CD037499 to CD038843) were designed by testing for polymorphic SSR markers among peanut genotypes. Nine of the 44 primer pairs resulted in polymorphism among the 24 genotypes. The number of alleles at each locus detected ranged from two to five. These EST-derived SSR markers would be useful in gene discovery and genetic mapping, and serve as anchor points in comparative mapping from different populations. The rate of detecting polymorphism among peanut lines is higher on the basis of EST-derived SSR markers (over 20%) than SSR derived from genomic sequences of peanut.

To identify novel genes, 400 unigenes have been selected from these two libraries and arrayed on glass slides for gene expression analyses. RT-PCR analysis was conducted to validate the microarray and EST data (Bustin, 2002) and to detect the respective expression

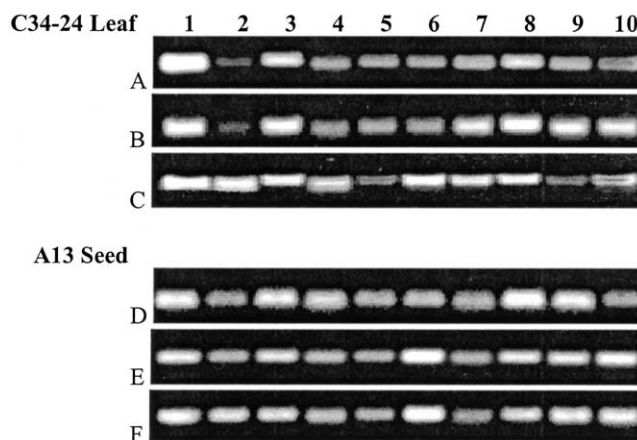


Fig. 2. Quantitative expression analysis of genes with putative function of disease resistance and drought tolerance by reverse transcription PCR (RT-PCR). Total RNA was extracted from peanut C34-24 leaf samples of healthy plants (A), drought stressed plants (B), or plants challenged by fungal leaf spots (C), and from peanut A13 immature seeds without stress (D), drought stressed (E), or challenged by fungus *Aspergillus flavus* under drought stress (F). Genes examined: 1-Histone H3.3 (positive internal control), 2-Allergen Ara h3/4, 3-10-kDa protein, 4-Leucine-rich protein, 5-Bax inhibitor, 6-Heat shock cognate protein, 7-Non-specific lipid transfer protein, 8-drought inducible 22-kDa protein, 9-drought inducible RPR 10, 10-Defensin protein.

pattern associated with different traits. Specific primers (Table 1) were designed for 10 specific genes from the two cDNA library EST sequences with high expression frequency. The primers recognize each specific gene, and no other products were amplified (Fig. 2). Several genes had relatively high expression in leaves or immature pods, such as nonspecific lipid-transfer protein in leaves, allergen Ara h3/Ara h4 in immature pods. Ara h3/Ara h4 was also expressed in the leaves infected with fungal pathogen but not in healthy leaf tissues. Several genes had no significant difference in expression, such as the 10-kDa protein precursor and the drought inducible 22-kDa protein. The relationship between putative resistant genes and the traits was not clear. However, genes such as allergen Ara h3/Ara h4, leucine-rich receptor-like protein kinase, and cell-autonomous heat shock cognate protein were expressed higher in the plants challenged with leaf spot pathogen, and cell-autonomous heat shock cognate protein and defensin protein were expressed higher in pods of plants challenged with *A. flavus* under drought stressed conditions.

During crop development and growth in the field, plants are challenged with different abiotic and biotic stresses such as drought, high salt concentration, cold, and diseases. To understand the response of plants to these stresses, researchers have studied plant stress responses for decades (Cushman and Bohnert, 2000). Genetic response to drought stress has been shown to be controlled by QTLs (quantitative trait loci) in drought-tolerant cultivars of rice (Quarrie et al., 1997). The information in genomic research could help geneticists and breeders identify genes expressed in a particular tissue or under a specific condition and could aid in understanding the complexity of gene expression and regulation.

CONCLUSIONS

The EST data from this study have resulted in the identification of many genes previously reported in other plants as responsive to adversity stresses. The results of the EST analysis described here indicate that it is a useful approach for isolating genes of *A. hypogaea* on the basis of homologous sequence comparison and for novel genes not previously isolated that are expected to be peanut-specific genes. Certainly, the homologous genes may not have the same function as they have in other plants, but the homology should be useful in formulating predictions about their functions. Many ESTs in these two cDNA libraries (matching stress-related genes, disease-related genes and allergen genes) imply that they are strongly conserved across plant families (Bennetzen, 2000). The number of ESTs and other genomic DNA sequences is increasing rapidly, and comparative genome studies may help identify useful genes in species with a large genome, such as *A. hypogaea* which has a genome of 2800 Mbp (Arumuganathan and Earle, 1991). The data present in this study showed many novel genes through BLAST search. Redundancy also exists as expected. Our next step is to characterize the function of these novel genes. Some of the techniques that can be utilized include subtractive cDNA library construction, microarrays in high-throughput hybridization screening, and bioinformatics in the analysis of gene expression profiling and data mining (Luo et al., 2002, 2003a, 2003b). Unfortunately, marker development in peanuts is limited, but EST-derived SSR markers could enhance marker identification, or genetic mapping. Because many agronomic traits are very difficult to select in peanut breeding by conventional selection techniques, marker-assisted selection offers an additional tool for genetic improvement of germplasm lines.

ACKNOWLEDGMENTS

We thank Jerry Mozoruk for technical help and Kayla Young for data input. This research is supported by funds provided by USDA Agricultural Research Service and National Peanut Foundation, and by USDA Specific Cooperative Agreement 58-6602-213. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

REFERENCES

- Anderson, W.F., C.C. Holbrook, and D.M. Wilson. 1996. Development of greenhouse screening for resistance to *Aspergillus parasiticus* infection and preharvest aflatoxin contamination in peanut. *Mycopathologia* 135:115-118.
- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Arumuganathan, K., and E.D. Earle. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9:208-218.
- Barre, A., Y. Bourne, E.J. Van Damme, W.J. Peumans, and P. Rouge. 2001. Mannose-binding plant lectins: Different structural scaffolds for a common sugar-recognition process. *Biochimie* 83:645-651.
- Bennetzen, J.L. 2000. Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. *Plant Cell* 12:1021-1029.
- Boguski, M.S., T.M. Lowe, and C.M. Tolstoshev. 1993. dbEST-database for "expressed sequence tags". *Nat. Genet.* 4:332-333.
- Boote, K.J. 1982. Growth stages of peanut (*Arachis hypogaea* L.). *Peanut Sci.* 9:35-40.
- Branch, W.D., and A.K. Culbreath. 1995. Combination of early maturity and leaf spot tolerance within an advanced Georgia peanut breeding line. *Peanut Sci.* 22:106-108.
- Bustin, S.A. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. *J. Mol. Endocrinol.* 29:23-39.
- Chapple, C. 1998. Molecular genetics analysis of plant cytochrome P450-dependent monooxygenases. *Annu. Rev. Plant Physiol. Mol. Biol.* 49:311-343.
- Cordeiro, G.M., R. Casu, C.L. McIntyre, J.M. Manners, and R.J. Henry. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* 160:1115-1123.
- Culbreath, A.K., J.W. Todd, D.W. Gorbet, S.L. Brown, J.A. Baldwin, H.R. Pappu, C.C. Holbrook, and F.M. Shokes. 1999. Response of early, medium, and late maturing peanut breeding lines to field epidemics of tomato spotted wilt. *Peanut Sci.* 26:100-106.
- Cushman, J.C., and H.J. Bohnert. 2000. Genomic approaches to plant stress tolerance. *Curr. Opin. Plant Biol.* 3:117-124.
- Gai, X., S. Lal, L. Xing, V. Brendel, and V. Walbot. 2000. Gene discovery using the maize genome database ZmDB. *Nucleic Acids Res.* 28:94-96.
- He, G., R. Meng, M. Newman, G. Gao, R.N. Pittman, and C.S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 3:3.
- Hellmann, H., and M. Estelle. 2002. Plant development: Regulation by protein degradation. *Science* 297:793-797.
- Holbrook, C.C., K.E. Matheron, D.M. Wilson, W.F. Anderson, M.E. Will, and A.J. Norden. 1994. Development of a large-scale field system for screening peanut for resistance to preharvest aflatoxin contamination. *Peanut Sci.* 21:20-22.
- Holbrook, C.C., and H.T. Stalker. 2003. Peanut breeding and genetic resources. *Plant Breed. Rev.* 22:297-356.
- Hopkins, M.S., A.M. Casa, T. Wang, S.E. Mitchell, R.E. Dean, G.D. Kochert, and S. Kresovich. 1999. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in cultivated peanut. *Crop Sci.* 39:1243-1247.
- Kantety, R.V., M. La Rota, D.E. Matthews, and M.E. Sorrells. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48:501-510.
- Knauff, D.A., A.J. Chiyembekeza, and D.W. Gorbet. 1992. Possible reproductive factors contributing to outcrossing in peanut. *Peanut Sci.* 19:29-31.
- Krapovickas, A., and W.C. Gregory. 1994. Taxonomy of the genus *Arachis* (Leguminosae). *Bonplandia* 8:1-186.
- Lazo, G.R., J. Tong, R. Miller, C. Hsia, C. Rausch, Y. Kang, and O.D. Anderson. 2001. Software scripts for quality checking of high-throughput nucleic acid sequencers. *Biotechniques* 30:1300-1305.
- Li, X.M., D. Serbrisky, S.Y. Lee, C.K. Huang, L. Bardina, B.H. Schofield, J.S. Stanley, A.W. Burks, G.A. Bannon, and H.A. Sampson. 2000. A murine model of peanut anaphylaxis: T- and B-cell responses to a major peanut allergen mimic human responses. *J. Allergy Immunol.* 106:150-158.
- Luo, M., X.Y. Kong, N.X. Huo, R.H. Zhou, and J.Z. Jia. 2002. EST analysis of resistance to powdery mildew disease in wheat at primary infected stage. *Acta Genet.* 29:525-530.
- Luo, M., N. Huo, C. Xu, X. Kong, and J.Z. Jia. 2003a. SSH cDNA library construction and macroarray preparation from wheat challenged by *Erysiphe graminis* DC. *Sci. Agric. Sinica* 36:49-53.
- Luo, M., P. Dang, B.Z. Guo, C.C. Holbrook, and M. Bausher. 2003b. Application of EST technology in functional genomics of *Arachis hypogaea* L. *Phytopathology* 94:S55.
- Martienssen, R.A. 2000. Weeding out the genes: The *Arabidopsis* genome project. *Funct. Integr. Genomics* 1:2-11.
- Mewes, H.W., D. Frishman, U. Guldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Munsterkotter, S. Rudd, and B. Weil. 2002. MIPS: A database for genomes and protein sequences. *Nucleic Acids Res.* 30:31-34.
- Mills, E.N., J. Jenkins, N. Marigheto, P.S. Belton, A.P. Gunning, and V.J. Morris. 2001. Allergens of the cupin superfamily. *Biochem. Soc. Trans.* 30:925-929.

- Miziorko, H.M., and G.H. Lorimer. 1982. Ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Annu. Rev. Biochem.* 52:507–535.
- Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yano, C.R. Bhatia, and T. Saski. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* 3:87–103.
- Nurnberger, T., and D. Scheel. 2001. Signal transmission in the plant immune response. *Trends Plant Sci.* 6:372–379.
- Ozias-Akins, P., and R. Gill. 2001. Progress in the development of tissue culture and transformation methods applicable to the production of transgenic peanut. *Peanut Sci.* 28:123–131.
- Park, C.J., R. Shin, J.M. Park, G.J. Lee, J.S. You, and K.H. Paek. 2002. Induction of pepper cDNA encoding a lipid transfer protein during the resistance response to tobacco mosaic virus. *Plant Mol. Biol.* 48:243–254.
- Pearce, R.S., C.E. Houlston, K.M. Atherton, J.E. Rixon, P. Harrison, M.A. Hughes, and A.M. Dunn. 1998. Localization of expression of three cold-induced genes, *b1t101*, *b1t4.9*, and *b1t14*, in different tissues of the crown and developing leaves of cold-acclimated cultivated barley. *Plant Physiol.* 117:787–795.
- Pons, L., C. Chery, A. Romano, F. Namour, M.C. Artesani, and J.L. Gueant. 2002. The 18 kDa peanut oleosin is a candidate allergen for IgE-mediated reactions to peanuts. *Allergy* 72:88–93.
- Quarrie, S.A., D.A. Laurie, J. Zhu, C. Lebreton, A. Semikhodskii, A. Steed, H. Witsenboer, and C. Calestani. 1997. QTL analysis to study the association between leaf size and abscisic acid accumulation in droughted rice leaves and comparisons across cereals. *Plant Mol. Biol.* 35:155–165.
- Richard, A.D., J.L. Chris, M. Sameer, J.H.S. Vincent, and L.P. Nancy. 1996. Metabolic engineering: Prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses—A review. *Gene* 179:61–71.
- Rozen, S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. p. 365–386. *In* S. Krawetz, and S. Misener (ed.) *Bioinformatics methods and protocols: Methods in molecular biology*. Humana Press, Totowa, NJ.
- Schoof, H., P. Zaccaria, H. Gundlach, K. Lemcke, S. Rudd, G. Kolesov, R. Arnold, H.W. Mewes, and K.F. Mayer. 2002. MIPS *Arabidopsis thaliana* Database (MAtdB): An integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res.* 30:91–93.
- Schuler, G.D. 1997. Pieces of the puzzle: Expressed sequence tags and the catalog of human genes. *J. Mol. Med.* 75:694–698.
- Shoemaker, R., P. Keim, L. Vodkin, E. Retzel, S.W. Clifton, R. Waterson, D. Smoller, V. Coryell, A. Khanna, and J. Erpelding. 2002. A compilation of soybean ESTs: Generation and analysis. *Genome* 45:329–338.
- Somerville, C., and S. Somerville. 1999. Plant functional genomics. *Science* 285:380–383.
- Stalker, H.T. 1997. Peanut (*Arachis hypogaea* L.). *Field Crops Res.* 53:205–217.
- Stalker, H.T., and L.G. Mazingo. 2001. Molecular markers of *Arachis* and marker-assisted selection. *Peanut Sci.* 28:117–123.
- Sun, W., M. van Montagu, and N. Verbruggen. 2002. Small heat shock proteins and stress tolerance in plants. *Biochim. Biophys. Acta* 1577:1–9.
- Terryn, N., P. Rouze, and M. van Montagu. 1999. Plant genomics. *FEBS Lett.* 452:3–6.
- Trevino, M.B., and M.A. O'Connell. 1998. Three drought-responsive members of the nonspecific lipid-transfer protein gene family in *Lycopersicon pennellii* show different developmental patterns of expression. *Plant Physiol.* 116:1461–1468.
- Wheeler, D.L., D.M. Church, S. Federhen, A.E. Lash, T.L. Madden, J.U. Pontius, G.D. Schuler, L.M. Schriml, E. Sequeira, T.A. Tatusova, and L. Wagner. 2003. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 31: 28–33.
- Wynne, J.C., and T.A. Coffelt. 1982. Genetics of *Arachis hypogaea* L. p. 50–94. *In* H.E. Pattee and C.T. Young (ed.) *Peanut science and technology*. Am. Peanut Res. Educ. Soc. Inc, Publishers, Texas.
- Xiang, P., T.A. Beardslee, M.G. Zeece, J. Markwell, and G. Sarath. 2002. Identification and analysis of a conserved immunoglobulin E-binding epitope in soybean G1a and G2a and peanut Ara h 3 glycinins. *Arch. Biochem. Biophys.* 408:51–57.
- Yamamoto, K., and T. Sasaki. 1997. Large-scale EST sequencing in rice. *Plant Mol. Biol.* 35:135–144.
- Ye, X.Y., and T.B. Ng. 2002. A new antifungal peptide from rice beans. *J. Pept. Res.* 60:81–87.
- Yu, J., et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296:79–92.
- Zhang, S., M.S. Reddy, N. Kokalis-Burelle, L.W. Wells, S.P. Nightengale, and J.W. Kloepper. 2001. Lack of induced systemic resistance in peanut to late leaf spot disease by plant growth-promoting rhizobacteria and chemical elicitors. *Plant Dis.* 85:879–884.
- Zhu, J.K. 2002. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53:247–273.