

Expressed sequence tags in cultivated peanut (*Arachis hypogaea*): discovery of genes in seed development and response to *Ralstonia solanacearum* challenge

Jiaquan Huang · Liying Yan · Yong Lei ·
Huifang Jiang · Xiaoping Ren · Boshou Liao

Received: 8 December 2011 / Accepted: 25 March 2012 / Published online: 31 May 2012
© The Botanical Society of Japan and Springer 2012

Abstract Although an important oil crop, peanut has only 162,030 expressed sequence tags (ESTs) publicly available, 86,943 of which are from cultivated plants. More ESTs from cultivated peanuts are needed for isolation of stress-resistant, tissue-specific and developmentally important genes. Here, we generated 63,234 ESTs from our 5 constructed peanut cDNA libraries of *Ralstonia solanacearum* challenged roots, *R. solanacearum* challenged leaves, and unchallenged cultured peanut roots, leaves and developing seeds. Among these ESTs, there were 14,547 unique sequences with 7,961 tentative consensus sequences and 6,586 singletons. Putative functions for 47.8 % of the sequences were identified, including transcription factors, tissue-specific genes, genes involved in fatty acid biosynthesis and oil formation regulation, and resistance gene analogue genes. Additionally, differentially expressed genes, including those involved in ethylene and jasmonic acid signal transduction pathways, from both peanut leaves and roots, were identified in *R. solanacearum* challenged samples. This large expression dataset from different peanut tissues will be a valuable source for marker development and gene expression analysis. It will also be helpful for finding candidate genes for fatty acid synthesis and oil formation regulation as well as for studying mechanisms of

interactions between the peanut host and *R. solanacearum* pathogen.

Keywords *Arachis hypogaea* L. · Expressed sequence tags · Fatty acid biosynthesis · Bacterial wilt resistance · Signal transduction

Introduction

Groundnut or peanut (*Arachis hypogaea* L.) is an important crop for oil and protein production. It provides one of the most valuable resources of nutrition for human consumption in developing countries. In China, the peanut planting area is about 5 million ha, with the total annual production of 14 million tons, and comprises 40 % of the world's peanut shell production (<http://faostat.fao.org/site/339/default.aspx>). Due to the fact that about 60 % of edible oil in China is imported, groundnut is a promising crop for oil production to meet the demand of domestic consumption. However, the oil content and fatty acid composition of most cultivated peanut plants need to be improved, as well as their resistance to biotic and abiotic stresses, which greatly affect peanut production.

Isolation and characterization of candidate genes is of vital importance for peanut genetic improvement. Although peanuts are morphologically diverse, and resistance to several pathogens (Guo et al. 2008) has been reported, the genetic diversity of cultivated peanut is relatively low (Kochert et al. 1996). As a consequence, polymorphic genetic markers are lacking in cultivated peanut. Currently, it is nearly impossible to identify candidate genes via the map-based cloning method. A variety of other techniques have been developed successfully to identify candidate genes in peanut, including two-dimensional electrophoresis

Electronic supplementary material The online version of this article (doi:10.1007/s10265-012-0491-9) contains supplementary material, which is available to authorized users.

J. Huang · L. Yan · Y. Lei · H. Jiang · X. Ren · B. Liao (✉)
Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China
e-mail: peanutresearch@hotmail.com

followed by tandem mass spectrometry of isolated proteins (Jain et al. 2006), suppression subtractive hybridization (Nobile et al. 2008; Tirumalaraju et al. 2011) and sequence homolog based cloning (Chi et al. 2010; Li et al. 2005). However, these gene isolation methods are based on sequence similarities between amino acid or nucleotide sequences of peanut and other sources. Expressed sequence tag (EST) profiling is also based on sequence similarity but provides an efficient means for gene isolation. It has been widely used to identify genes involved in specific biological functions, especially in organisms where genomic data are not available (Gueguen et al. 2003). ESTs can also be used for functional marker development, and they can provide a valid and reliable source of gene expression data. In an unbiased cDNA library, the number of ESTs matching particular genes should reflect the abundance of the corresponding cDNAs in the library and the level of their mRNAs in the tissue from which the library was derived (Ewing et al. 1999). EST sequencing has been widely used for gene isolation, and the public EST data has grown exponentially since its first application. In plants, this method was initially used for the model species *Arabidopsis thaliana* (Rounsley et al. 1996) and *Oryza sativa* (Yamamoto and Sasaki 1997), and has subsequently been used in many other plant species (Asamizu et al. 2000; Fernandes et al. 2002; Shoemaker et al. 2002; Sterky et al. 1998; Yamamoto and Sasaki 1997). With accumulating data from various plant species and functional analysis of new genes, the information based on sequence similarity will become more reliable and will greatly facilitate the study of peanut genes.

Compared with the large number of ESTs sequenced in other plants, such as *Arabidopsis*, rice, maize and wheat, peanut ESTs are still lacking. Before the submission of our peanut EST data, only about 162,030 *Arachis* ESTs have been deposited in Genbank and are publicly available. Cultivated peanut is an allopolyploid with a large genome of about 2,800 Mbp. Thus, compared with its large genome, the number of ESTs of peanut available is still small, which greatly hampers the advance of peanut genomic studies. Although peanut ESTs have increased significantly in recent years, and the data has been widely used for candidate gene isolation and marker development (Liang et al. 2009; Luo et al. 2005b), most of the data has been derived from peanut leaves and developing seeds. To our knowledge, ESTs from roots of cultivated peanut have not been sequenced. The gene expression patterns to biotic stresses such as leaf spot caused by *Cercosporidium personatum* (Luo et al. 2005a), virus infection caused by tomato spotted wilt virus (Guo et al. 2009) and aflatoxin contamination caused by *Aspergillus parasiticus* (Guo et al. 2008) have been analyzed via EST sequencing, but no information is available on how the host peanut plant

responds to *Ralstonia solanacearum* infection. Bacterial wilt in peanut caused by *R. solanacearum* is a serious problem which has been long prevalent in southern part of China (Liao 2005). In this study, a large set of expressed genes was sequenced from root, leaf and developing seed, which included both previously described and novel genes. This information will facilitate future genomic studies in peanut.

Materials and methods

Peanut cDNA library construction and quality assessment

Peanut line 06-4104 bearing the characteristics of high oil content and high resistance to bacterial wilt was used in our experiment. The oil content in the seeds reached 56 %, and the survival rate to the infectious agent *R. solanacearum* amounted to 95 % in three successive years in a field trial. Developing seeds were collected from R2 to R7 stages (Boote 1982) and pooled together to make the NSL cDNA library. Unchallenged leaves and roots were collected in peanut plants 20 days after sowing to make the NLL and NRL cDNA libraries, respectively. *R. solanacearum* challenged leaves and roots were also collected after 12, 24, 36, 48 and 72 h of infection, and the material from challenged roots and leaves were pooled to make the RLL and RRL cDNA libraries, respectively. Altogether, 5 cDNA libraries were constructed in our study.

Total RNA was extracted using Trizol agent (Invitrogen, USA) according to the manufacturer's instructions. And mRNA was subsequently purified using the mRNA PolyAtract@mRNA Isolation System (Promega, USA). The first and second strand cDNA was synthesized using the Creator SMART cDNA Library Construction Kit (from Clontech, Takara, Beijing, China) as suggested by the manufacturer. Double stranded cDNA was then subjected to *Sfi*I digestion, and fragments larger than 500 bp were collected and ligated with the pDNR-LIB vector.

The quality of the cDNA library was first assessed by colony PCR of 96 randomly picked clones to determine the average insert size and percentage of clones without inserts. Subsequently, 384 randomly picked clones were sequenced to determine the ratio of contaminating sequences (e.g., vector, *E. coli*) and average length of valid sequences. The randomness of the cDNA library was determined by calculating the ratio of unique sequences (contigs + singletons)/reads. The primary cDNA libraries with a titer more than 1×10^6 pfu/ml, recombination rate more than 99 % and the average inserts length of about 1.2 kb were used for subsequent analysis.

Plasmid preparation and 5' sequencing

Plasmid constructs were transformed into *E. coli* DB 10B (Invitrogen), grown overnight on solid LB containing IPTG and X-Gal. Approximately 10,000 clones were picked, grown overnight in LB medium containing chloramphenicol in 96-well plates. Plasmids were isolated and sequenced on an ABI 3730 sequencer using the M13 universal primer that anneals to the plasmid backbone upstream of the 5' end of the cDNA inserts.

Clustering of cDNA sequences

Phred (Ewing et al. 1998, 1999) was used with default parameters to determine each base call from files extracted from sequencers. Low quality sequences were rejected, and then the vector sequence was trimmed using a crossmatch program (Green 2004). DNA sequences of more than 100 bp were then accepted and saved in FASTA format. The EST contig assemble is done using the PHRAP program.

Functional annotation and classification

Unigenes were compared with the NCBI non-redundant nucleotide database and non-redundant protein database using BLASTN and BLASTX (Altschul et al. 1997), respectively, with the same *E* value cutoff of $1\text{E}-5$. Unigenes were identified by sequence similarity comparison against the SWISS-PROT database (SWISS-PROT was downloaded from the European Bioinformatics Institute) with BLAST at *E* values $\leq 1\text{E}-10$. Unigenes were assigned functional annotations by sequence similarity comparison against the Clusters of Orthologous Groups (COG) of proteins database (Tatusov et al. 1997, 2003) with BLAST at *E* values $\leq 1\text{E}-10$. A Perl script was written to assign the functional class of unigenes. Unigenes were first compared with the Kyoto Encyclopedia of Genes and Genomes (KEGG, release 46) database (Kanehisa et al. 2006) using BLASTX at *E* values $\leq 1\text{E}-10$. A Perl script was used to retrieve GO information from the BLAST result and to establish unigene pathway associated with the established database. InterPro domains (Mulder et al. 2003) were

annotated using InterProScan (Zdobnov and Apweiler 2001) release 16.0, and functional assignments were mapped using Gene Ontology (GO) (Harris et al. 2004). And WEGO (Ye et al. 2006) was used to carry out GO classifications and to generate a GO tree.

Results

After trimming low-quality sequence and removing vector and contaminant host sequences, a total number of 63,234 high-quality ESTs out of 68,788 clones were obtained with a mean length of 509 bp (Supplement figure 1). The ESTs were assembled into 7,961 contigs and 6,586 singletons, which altogether represented up to 7,346,235 bp and accounted for $\sim 0.3\%$ of the peanut genome (~ 2.8 Gbp). The developing seeds produced the most abundant unigenes, which comprised more than 50 % of the total unigenes, while unchallenged leaves produced the least number of genes ($<20\%$). It is likely that a small number of genes are needed for the unchallenged leaf to carry out its function, while more genes are required to play a biochemical roles during seed development and under pathogenic challenge attack, as has been demonstrated that only a small number of unigenes were isolated from leaves compared with that of roots and developing seeds (Table 1). The distribution of putative open reading frames (ORFs) of these unigenes indicated that the average length exceeded 300 bp (Supplement figure 1), which was useful for gene function annotation. The fact that the number of ESTs in each cluster ranged from 1 to 781 with an average of 4.35 per cluster (Supplement table 1) and all ESTs were obtained from non-normalized cDNA libraries, indicating that the data could be used to reveal global gene expression pattern.

Similarity and functional category analysis of peanut ESTs

Before the annotation process, the longest ORFs of all unigenes were determined by dynamically translating them in all six reading frames. A total of 363 unigenes containing short ORFs (≤ 100 nucleotide bases) were

Table 1 Assembly result of peanut cDNA libraries

Libraries	Assembled ESTs	Contigs	Singlets	Unigenes	Redundancy (%)
DSL	20,580	3,516	3,912	7,428	63.9
NLL	10,593	968	1,904	2,872	72.9
RILL	11,158	1,478	2,617	4,095	63.3
NRL	10,820	1,517	3,124	4,641	57.1
RIRL	10,056	1,511	3,183	4,694	53.3
Total	63,207	7,961	6,586	14,547	77.0

Table 2 Full-length evaluation of unigenes based on homology analysis

	Total	DSL	Leaf			Root		
			Total	NLL	RILL	Total	NRL	RIRL
Total unigenes	14,547	7,428	5,622	2,872	4,095	7,371	4,641	4,694
Match the homologue protein in <i>Arabidopsis thaliana</i> ^a	11,470	5,993	4,705	2,448	3,484	5,956	3,845	3,871
Match the head of homologue protein in <i>Arabidopsis thaliana</i>	4,850	2,752	2,167	1,127	1,719	2,812	1,866	1,981
Match the homologue protein in NCBI Nr ^b	11,498	6,040	4,668	2,421	3,466	5,968	3,855	3,852
Match the head of homologue protein in NCBI Nr	5,267	3,001	2,294	1,180	1,814	2,957	1,952	2,072

^a *Arabidopsis thaliana* protein data were downloaded from TAIR (<http://www.arabidopsis.org/>), June 19, 2009. Using the blast with *E* value $\leq 1\text{E}-10$ as cutoff

^b Data were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>), 20080310. Using the blast with *E* value $\leq 1\text{E}-10$ as cutoff

considered non-informative ones and were subsequently excluded from the following analysis. The remaining 14,184 ORF positive unigenes (containing 61,510 ESTs) were compared with the sequences deposited in NCBI nucleotide (Nt) database and NCBI non-redundant protein (Nr) database using BLASTN and BLASTX, respectively. Out of 13,932 high quality unigenes, 10,046 (72.1 %) had significant similarities with sequences in the Nt database, and 11,865 (85.2 %) had significant similarities with sequences in the Nr database (Table 2). The remaining 2,067 un-annotated unigenes may represent novel peanut-specific genes.

Sequences of peanut ESTs from the 5 libraries were further characterized by classification into functional categories by COG protein analysis (Tatusov et al. 1997, 2003). Details of the genes included in each category are given in Supplement figure 2. Only 3,317 (22.8 %) out of the total unigenes could be assigned to functional categories. Among them, metabolism occupied a significant proportion (39.4 %), followed by cellular processes and signalling (21.5 %), information storage and processing (20.4 %) and poorly characterized ones (18.7 %).

From the beginning of the experiment, we strived to obtain a high proportion of full-length genes in the cDNA libraries. Using the *Arabidopsis* genome as a reference, we found that ~ 42.3 % of the unigenes had a 5' head sequence (containing the authentic or putative start codon) via BLASTX analysis. However, when using the non-redundant database, about 45.8 % of unigenes had a 5' head sequence. The rates of genes with a 5' head sequence in each of the libraries are listed in Table 2. Because we used oligo dT as primer for reverse transcription, few of the genes did not contain the 3' end (poly A tail), as confirmed by clone isolation and sequencing. Full-length genes inserted in the plasmids will greatly facilitate their functional analysis, otherwise additional time and effort to performing rapid amplification of cDNA end (RACE) will be required.

Using BLASTN to search the Nt database, ~ 70.8 % of the unigenes could be annotated. Through BLASTX

analysis, 12,117 (83.3 %) showed significant similarity to the sequence in the Nr database, and the remaining 2,430 unigenes with no significant similarity were considered novel genes in *A. hypogaea*. Out of 14,547 unigenes analyzed via BLASTX, a total of 1,461 unigenes matched first to *Arabidopsis* sequences, which included 65 hypothetical proteins and 377 unknown or unnamed proteins. A total of 633 and 761 sequences matched first to soybean and medicargo, respectively, while only 192 unigenes matched first to the peanut sequence. Surprisingly, a large portion of the ESTs (5,730) matched first to *Vitis vinifera*, but only 6 of those sequences were functionally annotated. The remaining sequences were hypothetical or unnamed protein products, indicating that most of the genes expressed in peanut shared low sequence similarity with known and functionally analyzed sequences.

We were able to compare the unigene sequences with those of other species in the public genome databases. As expected, peanut ESTs showed the highest similarity to ESTs of soybean (mean identities of putative coding sequences (CDS) and proteins were 86.27 and 74.65 %, respectively), followed by another legume crop *Medicago truncatula*. The mean sequence similarities in the CDS and protein between peanut and the dicot *Arabidopsis* plant were 84.05 and 64.17 %, respectively. The least sequence similarities were detected in monocot cereal crops of rice and sorghum. The detailed comparisons of sequence similarities were illustrated in Supplement figure 3. The fact that similarities between peanut and other plants exceeded 84 % in CDS and 60 % in protein suggested that the peanut ESTs will be very valuable in that much genomic information can be extracted from the sequences by bioinformatic analysis.

Transcription factors in peanuts

Because transcription factors play important regulatory roles in genomes, we characterized related sequences in

peanut ESTs using Arabidopsis transcription factor datasets as queries. In peanut, a total of 3,505 ESTs encoding 1,062 putative transcription factors were identified. The most abundantly expressed transcription factors belonged to the Trihelix and C2H2 family, accounting for ~45 % of the total expressed transcription factors, followed by AP2-EREBP, C3H, bHLH, NAC, bZIP, WRKY, TCP, MYB, C2C2-CO-like, Homeobox, GRAS and CCAAT-HAP2 transcription factors, each of which was sequenced more than 50 times in this EST project (Fig. 1). These transcription factors were rarely expressed, with <10 clones sequenced in the respective cDNA libraries (data not shown). At the same time, most of these transcription factors belonged to gene families containing many unigenes, and the expression of each unigene was rare compared with the large number of all transcription factors. For example, among the most commonly found transcription factors sequences that contain C2H2 or the Trihelix domain, the majority of unigenes were represented by only one clone (Supplement figure 4).

Identification of tissue-specific genes in peanut

More than 10,000 clones in each cDNA libraries derived from leaves, roots and developing seeds were sequenced from the same peanut line, which allowed us to search for tissue-specific genes based on their expression pattern. When an EST is highly expressed (based on cluster size) in

one tissue but is not sequenced in other tissues, it is likely that the gene is tissue-specific. With larger numbers of sequenced ESTs in different libraries, the probability of determining such genes becomes higher. Table 3 lists detailed information on the highly expressed tissue-specific genes of peanut. Most of the seed-specific genes were storage protein related genes, such as those encoding arachin, allergen and globulin, as well as those related to fatty acid metabolism, such as oleosin and desaturase. In contrast, most genes with leaf-specific expression were related to photosynthesis. Although many genes were annotated to genomic or marker sequences of other plants based on BLAST analysis to Nt and Nr database, further annotation with the Swissprot and/or KEGG databases indicated that these genes were all related to photosynthesis. Similar to genes in leaves, many root-specific genes were annotated as genomic or marker sequences according to their sequence similarity. When comparing the seed- and leaf-specific genes, the abundance of root-specific genes was much lower, accounting for only 0.5 % of the total genes expressed.

Of the 14,547 unigenes clustered from peanut leaf, root and seed, 4,188 genes were specifically expressed in the developing seed, 3,301 genes were specifically expressed in root, and 2,298 genes were expressed only in leaf (Fig. 2). Meanwhile, 1,114 genes were expressed in all tissues, indicating that they were constitutively expressed genes.

Fig. 1 Number of unigenes in each transcription factor family. Transcription factors of Arabidopsis were searched against *A. hypogaea* ESTs, and then the target genes were classified according to the sequence similarity with the known transcription factor family. Each bar represents the number of unigenes in one transcription factor family

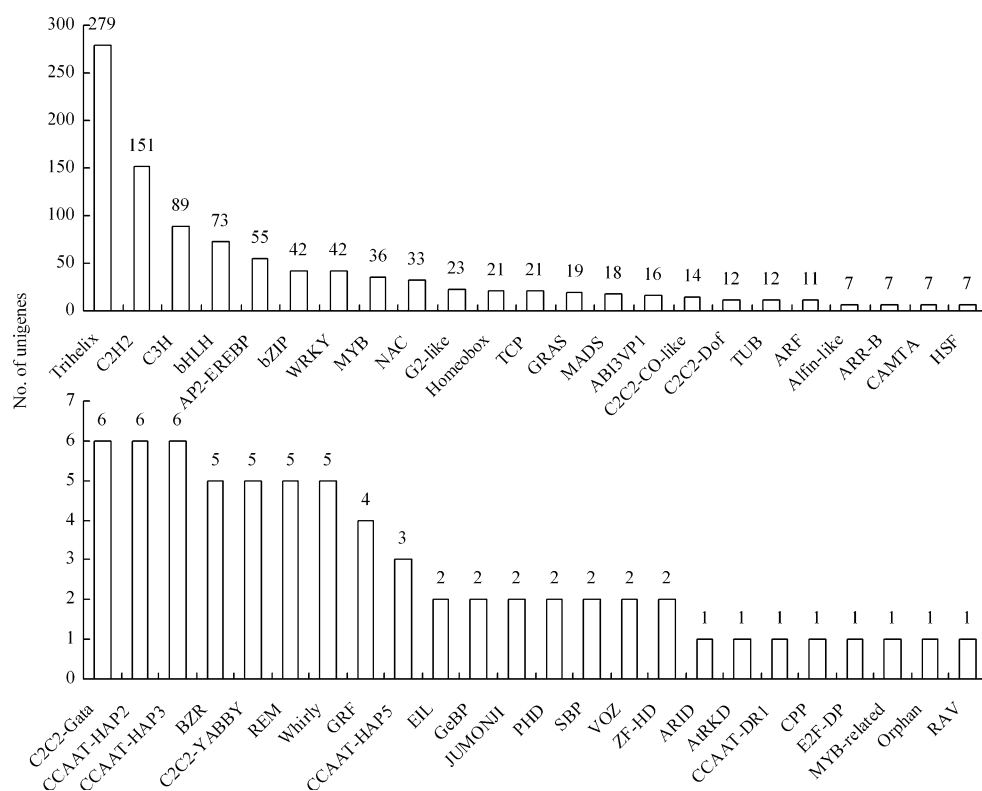


Table 3 Abundantly expressed tissue specific ESTs in peanut

Unigene ID	Accession no.	E value	Annotation	Organism	No. of clones sequenced			
					DSL	NLL	RILL	RURL
Seed specifically expressed genes								
OCPeontig12738-13	gb AAG01363.1	1.00E-116	Arachin	Arachis hypogaea	1,277	0	0	0
OCPeontig1273-5	sp P43238	1.00E-142	Allergen Ara h 1	Arachis hypogaea	1,185	0	0	0
OCPeontig12712	gb AAU21499.2	3.00E-67	Oleosin 1	Arachis hypogaea	116	0	0	0
OCPeontig12555	gb ABQ62972.1	2.00E-14	TINY-like protein	Populus trichocarpa	33	0	0	0
OCPeontig12575	gb AB184252.1	6.00E-82	Tonoplast intrinsic protein alpha TIP	Arachis hypogaea	35	0	0	0
OCPeontig12520	gb ABK92180.1	6.00E-89	11S seed storage globulin precursor B1	Arachis hypogaea	30	0	0	0
OCPeontig12425	gb AAF82293.1	1.00E-107	Microsomal oleate desaturase	Arachis hypogaea	23	0	0	0
OCPeontig12366	sp Q04672	8.00E-44	Sucrose-binding protein precursor	Glycine max	21	0	0	0
OCPeontig12296	gb AAF61730.1	1.00E-69	Beta-ketoacyl-ACP synthetase I	Glycine max	18	0	0	0
OCPeontig12406-2	emb CAN73668.1	4.00E-99	Hypothetical protein	Vitis vinifera	17	0	0	0
Leaf specifically expressed genes								
OCPeontig12740	sp Q02980	3.00E-45	Ribulose biphosphate carboxylase	Vitis vinifera	0	432	281	0
OCPeontig12734-2	gb ABK95667.1	2.00E-08	Photosystem II subunit T	Arabidopsis thaliana	0	312	3	0
OCPeontig12725	gb AB184258.1	3.00E-56	Photosystem I psaH protein	Vitis vinifera	0	164	20	0
OCPeontig12723-2	gb ABC46708.1	1.00E-36	Chloroplast photosystem II 10 kDa protein	Vitis vinifera	0	135	39	0
OCPeontig12726	gb AAQ84168.1	9.00E-89	PSI reaction center subunit III	Citrus sinensis	0	122	54	0
OCPeontig12701-3	gb ABK96439.1	2.00E-23	Photosystem II subunit X	Populus trichocarpa	0	81	6	0
OCPeontig12711-7	db JBAF80474.1	9.00E-28	Photosystem I reaction center subunit	Vitis vinifera	0	73	21	0
OCPeontig12722-2	gb AAG61120.1	3.00E-70	Ribulose biphosphate carboxylase	Vitis vinifera	0	62	54	0
OCPeontig12695	gb ABK96164.1	2.00E-52	Plastocyanin, chloroplast precursor	Solanum lycopersicum	0	50	33	0
Root specifically expressed genes								
OCPeontig12677-2	-	-	Hypothetical protein	Vitis vinifera	0	0	0	18
OCPeontig12603	emb CAN61352.1	3.00E-22	Protease inhibitor	Vitis vinifera	0	0	0	8
OCPeontig12671-2	gb ACA23207.1	1.00E-44	Kunitz trypsin protease inhibitor	Glycine max	0	0	0	22
OCPeontig12617-2	emb CAB45651.1	2.00E-90	Plasma membrane intrinsic protein	Pisum sativum	0	0	0	15
OCPeontig12539	ref XP_001490174.1	2.00E-09	Keratin associated protein 12.2	Equus caballus	0	0	0	13
OCPeontig12476-2	gb AA02947.1	6.00E-84	Nodulin-26	Glycine max	0	0	0	5
OCPeontig12380	emb CAO69871.1	1.00E-50	Hypothetical protein	Vitis vinifera	0	0	0	5
OCPeontig12451-2	ref NP_567391.1	1.00E-22	Protease inhibitor	Arabidopsis thaliana	0	0	0	1
OCPeontig12453-4	emb CAO21718.1	1.00E-51	Transcription factor MYB59	Vitis vinifera	0	0	0	5
OCPeontig12429-4	gb AAG34807.1	7.00E-72	Glutathione S-transferase 17	Glycine max	0	0	0	0

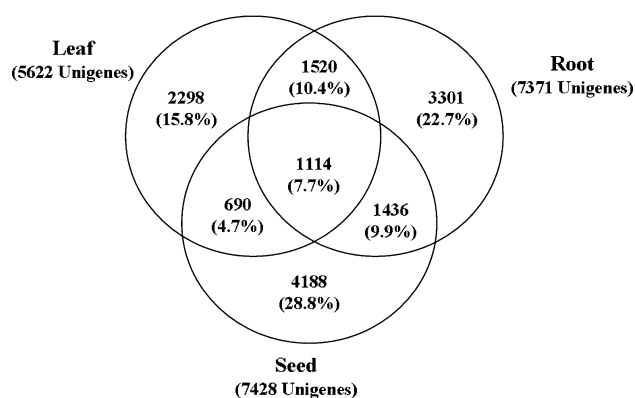


Fig. 2 Overlapping of unique EST sequences from different tissues of peanut

Isolation of fatty acid biosynthesis and oil formation genes in peanut

Because most genes for fatty acid biosynthesis and oil formation are well studied in plants such as Arabidopsis, soybean and rapeseed, it is possible to isolate corresponding genes in peanut based on nucleotide and putative amino acid sequence homology. As expected, most genes for fatty acid and oil formation in peanut have been sequenced, including previously functionally annotated *DGAT*, *Oleosin*, *FAD2*, *FAT*, *PEPC*. Some genes, such as *DGAT* and *FAD2*, have even been well characterized in peanut (Chu et al. 2009; Saha et al. 2006). Recently, those genes encoding acetyl-CoA carboxylases, acyl carrier protein, type II fatty acid synthase enzymes have also been analyzed in peanut (Li et al. 2009, 2010a, b). However, some genes involved in fatty acid biosynthesis in peanut have not been isolated before, and sequence search identified genes homologous to those from soybean, Arabidopsis, maize, rice, and even lotus and grape. Most genes were expressed in peanut seeds, while some were also expressed in root and leaf tissues, indicating that some fatty acid biosynthesis pathways play fundamental roles for the survival and growth of the peanut plant (Supplement table 2). We also isolated regulatory genes that may control oil content in peanut, such as *LEAFY COTYLEDON1*, *FUSCA*, *abscisic acid insensitive*, *dof*, *glabra2* and *phospholipase D*, which have been previously confirmed in Arabidopsis and other plants to play important roles in regulation of oil formation. Although these regulatory genes were low in abundance, the expression of most genes was predominantly in the seed, making them potential target genes for improvement of oil content and fatty acid composition.

Resistance gene analogues isolation from peanuts

Out of the 14,547 unigenes isolated from peanut, only 81 unigenes were putative resistance gene analogues (RGAs),

most RGAs were sequenced <4 times from all 5 cDNA libraries constructed, and the highest number of clones from one putative RGA was only 10. The low abundance of these putative RGAs in different peanut libraries made it difficult to correlate their expression with peanut bacterial wilt resistance. The RGAs predominantly expressed in peanut in this study were genes containing the TIR-NBS-LRR and CC-NBS-LRR domains (62 and 17 unigenes, respectively). Out of 81 RGAs analyzed, 6 RGAs were expressed in roots, leaves and developing seeds, indicating that they were constitutively expressed and could be involved in the basal resistance of plants. It is also interesting to note that of the 81 putative RGAs analyzed, 14 unigenes also belonged to the Trihelix transcriptional factor family, and 23 unigenes belonged to the WRKY transcriptional factor family (Supplement table 3), indicating that some disease resistance genes can lead to the expression of multiple functionally important genes.

Differentially expressed genes after *R. solanacearum* infection

Gene expression changes were detected in the highly resistant peanut line, and the results demonstrated that the patterns of expressed transcripts of both leaf and root were greatly changed after *R. solanacearum* infection. Of the total 5,622 leaf unigenes, 2,750 were newly expressed gene after pathogenic infection. Similar changes occurred in roots, and about 37 % of the total genes were novel (Fig. 3). The GO analysis categorized the functions of most of the differentially expressed genes to cell and cell part, binding and catalytic activity, and cellular and metabolic processes (Fig. 4). Similar expression patterns were detected in both leaf and root, indicating that these gene categories may be pathogen-related or responsible for disease resistance in peanut. Those genes involved in post-translational modification and translation and ribosomal structure were up-regulated after pathogenic challenge in both leaf and root, indicating a common signal transduction process occurred in both tissues and may have benefited the peanut plant in recognizing and defending against the pathogenic attack.

Possible signal transduction pathway for bacterial wilt resistance in peanut

We hypothesized that the specific genes stimulated by or involved in resistance to *R. solanacearum* attack should have similar signal transduction pathways in leaves and roots of peanut. To find the pathway that may participate in disease resistance, the criteria set for differential expression were less stringent, and the ratio of clones of two libraries >1.5 was considered to be significantly different in the

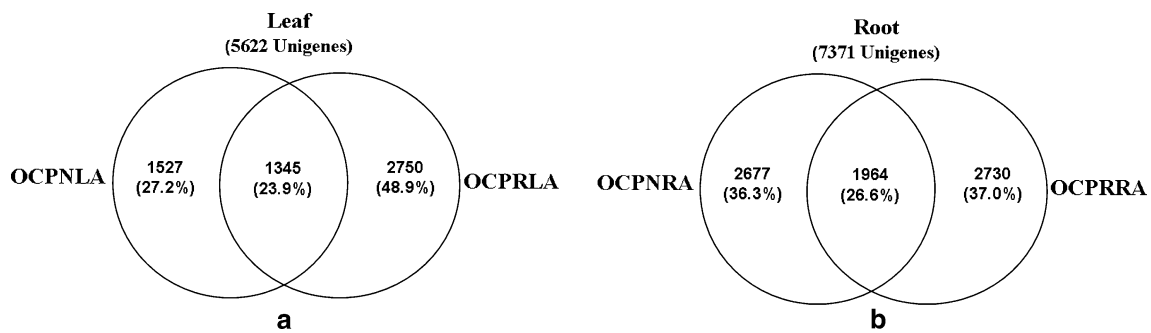


Fig. 3 Overlapping of peanut EST sequences. **a** Common and unique genes of unchallenged and *R. solanacearum* challenged peanut leaf. **b** Common and unique genes of unchallenged and *R. solanacearum* challenged peanut root. OCPNLA, OCPRLA, OCPNRA, OCPRRA

represents unchallenged leaf library, *R. solanacearum* challenged leaf library, unchallenged root library, *R. solanacearum* challenged root library, respectively

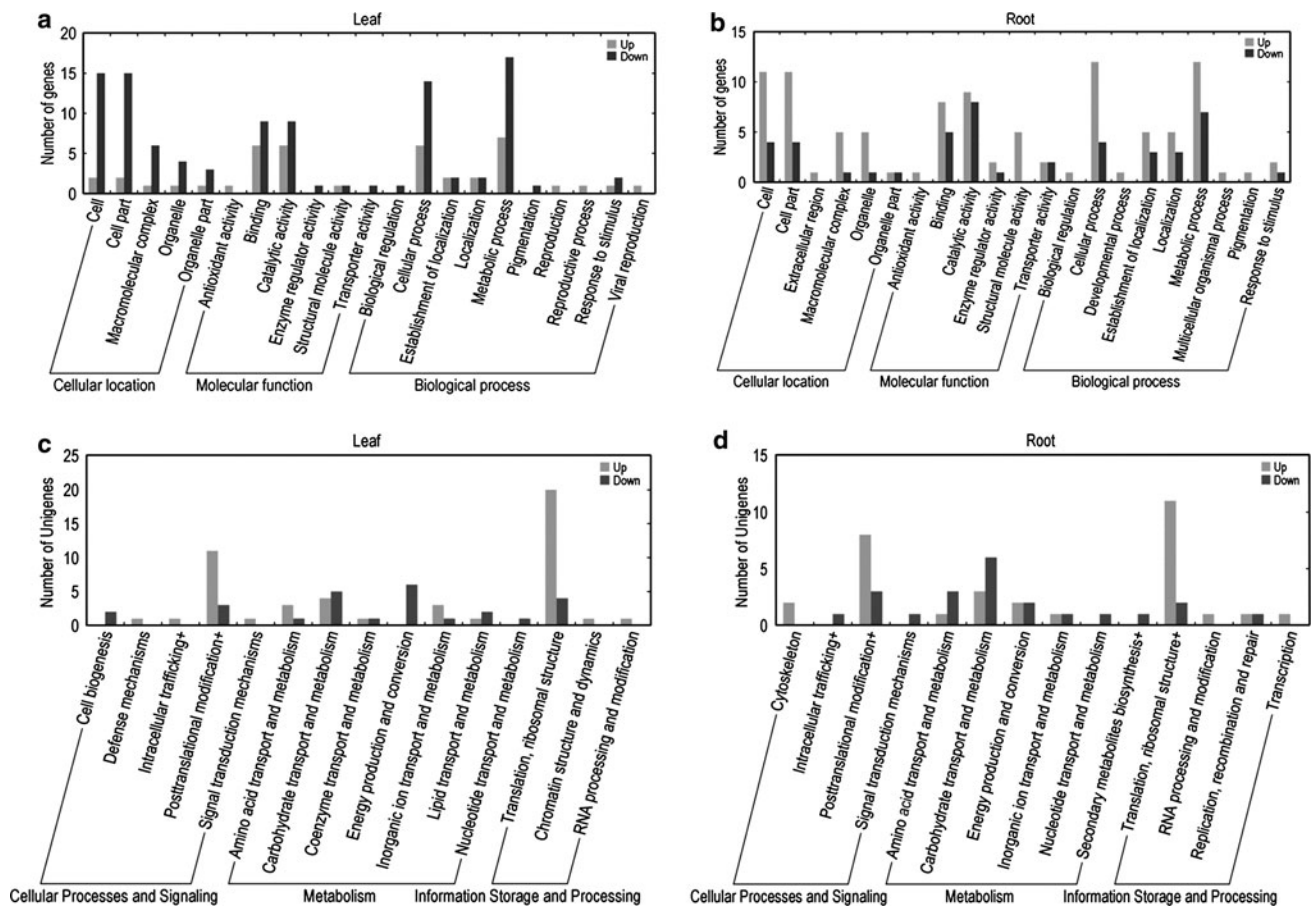


Fig. 4 GO and COG analysis of differentially expressed genes in peanut leaf and root after *R. solanacearum* challenge. **a** GO analysis of differentially expressed genes in peanut leaf after *R. solanacearum* challenge, **b** GO analysis of differentially expressed genes in peanut

root after *R. solanacearum* challenge, **c** COG analysis of differentially expressed genes in peanut leaf after *R. solanacearum* challenge, **d** COG analysis of differentially expressed genes in peanut root after *R. solanacearum* challenge

analysis. Although a large portion of the genes could not be characterized, many differentially expressed genes were ethylene responsive transcription factors or were involved in the ethylene signal transduction pathway (Table 4). Out of 51 genes encoding putative ethylene responsive transcription factors, 12 genes were up-regulated either in leaf

or root. Other components of the ethylene signal transduction pathway, such as *EIN3*, *MAPK*, *MAPKKK* and *EIL*, as well as genes involved in ethylene synthesis such as *ACC oxidase* were also expressed in both root and leave after pathogenic attack, indicating that ethylene biosynthesis or the ethylene signal transduction pathway may play

Table 4 Expression of ethylene signal transduction related genes after pathogen attack

Clone ID	Accession no.	Species	Gene description	E value	Expression	
					Leave	Root
Ethylene-responsive transcription factors						
OCPeontig8424	gblABO40237.1l	<i>Medicago truncatula</i>	ERF ^a 1A	5.00E−63	UP	UP
OCPeontig7942	refIXP_001763731.1l	<i>Arabidopsis thaliana</i>	ERF	1.00E−28	UP	ND
OCPeontig9386	gblAAX68525.1l	<i>Gossypium hirsutum</i>	Putative ERF ^a binding protein 2	2.00E−41	UP	ND
OCPeontig9898	embICAO71373.1l	<i>Arabidopsis thaliana</i>	ERF ^a 1B	5.00E−28	UP	UP
OCPeontig10844	gblABQ62991.1l	<i>Arabidopsis thaliana</i>	ERF ^a ERF008	4.00E−24	UP	ND
OCPeontig11843	gblAAQ96342.1l	<i>Vitis aestivalis</i>	Putative ERF ^a ERF3b	2.00E−21	ND	UP
OCPeontig11875	gblABX58051.1l	<i>Arabidopsis thaliana</i>	ERF ^a ERF010	2.00E−52	UP	UP
OCPeontig12056	gblABO40237.1l	<i>Medicago truncatula</i>	ERF ^a 1A	2.00E−41	UP	ND
OCPeontig12186	gblAAQ57226.1l	<i>Arabidopsis thaliana</i>	ERF ^a ERF060	9.00E−64	UP	UP
OCPeontig12237	refINP_196375.2l	<i>Arabidopsis thaliana</i>	ERF ^a ERF106	2.00E−31	UP	ND
OCPeontig12574	gblAAQ10777.1l	<i>Glycine max</i>	Ethylene responsive protein	4.00E−36	UP	UP
OCPeontig12627	gblABC69353.1l	<i>Nicotiana sylvestris</i>	ERF ^a 4	3.00E−21	UP	ND
OCPeontig10252	embICAO24070.1l	<i>Nicotiana tabacum</i>	ERF ^a 3	3.00E−30	DOWN	ND
EIN3						
OCPeontig6599	gblAAL76272.1l	<i>Arabidopsis thaliana</i>	Ethylene insensitive 3	1.00E−114	ND	UP
MAPK						
OCPeontig7630	dbjJBAE95414.1l	<i>Arabidopsis thaliana</i>	Mitogen-activated protein kinase 5	6.00E−38	UP	ND
OCPeontig7922	gblABR10070.1l	<i>Medicago sativa</i>	Mitogen-activated protein kinase homolog MMK2	5.00E−74	ND	UP
MAPKKK						
OCPeontig1488	embICAB16796.1l	<i>Arabidopsis thaliana</i>	MAP3K-like protein kinase	1.00E−33	ND	UP
CTR1						
OCPeontig4639	gblABN07953.1l	<i>Arabidopsis thaliana</i>	Serine/threonine-protein kinase CTR1	1.00E−100	UP	ND
ACC oxidase						
OCPeontig11275	embICAD70622.1l	<i>Cicer arietinum</i>	ACC	3.00E−73	ND	UP
OCPeontig9418	embICAO71363.1l	<i>Arabidopsis thaliana</i>	1-aminocyclopropane-1-carboxylate oxidase homolog 1	3.00E−40	UP	UP
OCPeontig11275	embICAD70622.1l	<i>Cicer arietinum</i>	1-aminocyclopropane-1-carboxylic acid oxidase	3.00E−73	ND	UP
EIL1						
OCPeontig6599	gblAAL76272.1l	<i>Vigna radiata</i>	transcription factor EIL1	1.00E−114	ND	UP
ACC deaminase						
OCPeontig12087	embICAO68552.1l	<i>Pyrococcus horikoshii</i>	1-aminocyclopropane-1-carboxylate deaminase (putative)	2.00E−67	ND	ND

UP, ND, DOWN represent up-regulated, non-differentially expressed and down-regulated genes, respectively, after *R. solanacearum* challenge. ERF^a represented ethylene-responsive transcription factor

a role in peanut bacterial wilt disease resistance. Similarly, many components of the jasmonic acid (JA) signal transduction pathway genes, such as *coil* and *cullin* were also significantly affected after pathogenic attack, and JA biosynthesis genes such as *LOX*, *AOS*, *AOC*, *OPR* and *JA carboxyl methyltransferase* were also significantly increased after bacterial infection (Table 5). These results indicated that resistance to bacterial wilt is probably associated with the ethylene and JA signal transduction pathways, and the detailed analysis of those genes may help to uncover the mechanism of bacterial wilt resistance in peanut. Calmodulin-related genes were also abundantly expressed in both root and leaf, but their expression levels did not greatly differed in root and leaf after pathogenic attack. At the same time, genes corresponding to the salicylic acid (SA) signal transduction pathway could hardly be detected in both root and leaf, indicating that this pathway is probably not involved in peanut bacterial wilt resistance or response.

Secondary metabolism may be involved in resistance to bacterial wilt

Secondary metabolism is an important plant defense system, but whether it is involved in peanut bacterial wilt resistance is still unknown. Our results indicated that many genes that did not change in leaf were significantly increased in roots after pathogenic attack, including

polyphenol oxidase, *phenylalanine ammonia lyase (PAL)*, many isoforms of *resveratrol synthases*, *isoflavone reductase* and *cytochrome P450 monooxygenase*. A small proportion of genes were up-regulated only in leaf such as those encoding thiazole biosynthetic enzyme, and even the expression of the gene encoding nitrilase was down-regulated after bacterial infection (Table 6). Considering that in the field, the first and the most critical step for bacterial wilt disease initiation and development is the interaction of pathogen and root of peanut, it is possible that the root is more sensitive than leaf. Therefore, the secondary metabolism of peanut in root was stimulated to a greater extent compared to that of leaf after pathogenic attack, indicating that the induction of genes involved in secondary metabolism likely contributes to increased resistance to bacterial wilt in peanut.

Discussion

Large-scale EST sequencing is a useful means to understand the genome of organisms due to the massive information provided by the genome-scale expression data. As of June 2011, ~162,000 *Arachis* ESTs were publicly available. Among these, 86,943 genes were from cultivated peanut (*A. hypogaea*), and 75,087 ESTs were from wild peanut species, *A. duranensis*, *A. ipaensis* and *A. stenosperma*. A number of EST sequencing projects has been

Table 5 Expression of jasmonic acid signal transduction related genes after pathogen attack

Genes	Accession no.	Species	Gene description	E value	Expression	
					Leave	Root
coil						
OCPcontig11208	gblAAZ66745.11	Arabidopsis thaliana	Coronatine-insensitive protein 1	8.00E−58	ND	UP
OCPcontig11333	gblAAZ66745.11	Arabidopsis thaliana	Coronatine-insensitive protein 1	1.00E−108	ND	UP
cullin						
OCPcontig9735	dbj1BAC10548.11	Arabidopsis thaliana	Cullin-1	5.00E−77	DOWN	ND
LOX						
OCPcontig6948	gblABS32276.11	Lens culinaris	Lipoxygenase	1.00E−121	UP	UP
OCPcontig7711	gblAAY87056.11	Glycine max	Seed lipoxygenase-2	6.00E−58	UP	UP
AOS						
OCPcontig4546	emblCAC86899.11	Glycine max	Allene oxide synthase	9.00E−39	ND	UP
AOC						
OCPcontig11463	gblABK94763.11	Arabidopsis thaliana	Allene oxide cyclase 3	5.00E−87	UP	UP
OPR						
OCPcontig11304	gblAAY27752.11	Arabidopsis thaliana	12-oxophytodienoate reductase 3	1.00E−136	UP	DOWN
OCPcontig786	dbj1BAA12160.11	Arabidopsis thaliana	12-oxophytodienoate reductase 2	2.00E−54	UP	ND
Jasmonic acid carboxyl methyltransferase						
OCPcontig8825	gblAAO27257.11	Brassica rapa	Jasmonate O-methyltransferase	3.00E−56	UP	ND

UP, ND, DOWN represent up-regulated, non-differentially expressed and down-regulated genes, respectively, after *R. solanacearum* challenge

Table 6 Expression of genes related to secondary metabolism after pathogen attack

Genes	Accession no.	Species	Gene description	E value	expression	
					Leave	Root
Polyphenol oxidase						
OCPeontig10587	gblABI96939.1l	<i>Malus domestica</i>	Polyphenol oxidase, chloroplast precursor	6.00E–23	ND	UP
OCPeontig11386	emblCAO23306.1l	<i>Malus domestica</i>	Polyphenol oxidase, chloroplast precursor	1.00E–41	ND	UP
Phenylalanine ammonia-lyase						
OCPeontig9501	splP45732l	<i>Stylosanthes humilis</i>	Phenylalanine ammonia-lyase	1.00E–79	ND	UP
Resveratrol synthase						
OCPeontig12119-2	gblABY86219.1l	<i>Arachis hypogaea</i>	Stilbene synthase 1	5.00E–94	ND	UP
OCPeontig12119-4	gblABY86219.1l	<i>Arachis hypogaea</i>	Stilbene synthase 1	5.00E–95	ND	UP
OCPeontig12666-7	splP51069l	<i>Arachis hypogaea</i>	Stilbene synthase 3	5.00E–94	ND	UP
OCPeontig12666-8	gblABY86219.1l	<i>Arachis hypogaea</i>	Stilbene synthase 1	8.00E–89	ND	UP
OCPeontig12666-10	gblAAF71253.1l	<i>Arachis hypogaea</i>	Stilbene synthase 1	1.00E–94	ND	UP
OCPeontig12666-11	splP51069l	<i>Arachis hypogaea</i>	Stilbene synthase 3	4.00E–89	ND	UP
OCPeontig12666-14	gblABY86219.1l	<i>Arachis hypogaea</i>	Stilbene synthase 1	2.00E–94	ND	UP
OCPeontig12666-15	splP51069l	<i>Arachis hypogaea</i>	Stilbene synthase 3	1.00E–136	ND	UP
OCPeontig12666-16	splP51069l	<i>Arachis hypogaea</i>	Stilbene synthase 3	4.00E–94	UP	UP
OCPeontig12666-17	gblAAX32891.1l	<i>Arachis hypogaea</i>	Stilbene synthase 3	4.00E–97	UP	UP
Leucoanthocyanidin dioxygenase						
OCPeontig9872	emblCAO50138.1l	<i>Arabidopsis thaliana</i>	Leucoanthocyanidin dioxygenase	3.00E–37	ND	UP
Isoflavone reductase						
OCPeontig11293	dbjJBAF34843.1l	<i>Solanum tuberosum</i>	Isoflavone reductase homolog	1.00E–67	ND	UP
Cytochrome P450 monooxygenase (CP450M)						
OCPeontig10613	gblABC59081.1l	<i>Medicago truncatula</i>	CP450M CYP92A29	1.00E–110	UP	UP
OCPeontig11309	gblABC68404.1l	<i>Glycine max</i>	CP450M CYP72A67	1.00E–46	ND	UP
OCPeontig11309-2	gblABC59088.1l	<i>Medicago truncatula</i>	CP450M CYP72A61	2.00E–89	ND	UP
Thiazole biosynthetic enzyme						
OCPeontig11287	gblABI84250.1l	<i>Citrus sinensis</i>	Thiazole biosynthetic enzyme, chloroplast precursor	1.00E–83	UP	ND
Nitrilase						
OCPeontig11000	gblAAT36331.1l	<i>Nicotiana tabacum</i>	Nitrilase 4	2.00E–65	ND	DOWN
OCPeontig11000-2	emblCAO43865.1l	<i>Rattus norvegicus</i>	Nitrilase homolog 2	5.00E–59	DOWN	DOWN

UP, ND, DOWN represent up-regulated, non-differentially expressed and down-regulated genes, respectively, after *R. solanacearum* challenge

carried out in peanut previously (Bi et al. 2010; Dang et al. 2008; Guo et al. 2008; Luo et al. 2005b; Yan et al. 2005). These data have helped to clarify nature of the resistance of peanut plants to stress such as fungus and virus infection, as well as to provide useful information for marker development. A genetic map based on EST-SSR markers has been developed (Liang et al. 2009). By analyzing 63,207 high-quality ESTs and the clustered 14,547 unigenes from developing seeds, roots and leaves in this study, we were able to draw an overall picture of its profiles and to gain a better understanding of the genes encoded by the peanut genome. The expression profile will be useful in studying molecular functions during peanut development and in discovering potential bacterial wilt resistance and agronomically important genes. Although much information is still lacking because many genes could not be functionally annotated or because bioinformatics analysis based on sequence comparisons would require experimental validation, the accumulated data produced here and those already deposited in GenBank will aid in developing high-density cDNA microarrays for transcriptome analysis to find target genes. The ESTs of the same gene from various peanut varieties will also help to find new single nucleotide polymorphic (SNP) markers, which would greatly improve the efficiency of marker selection because of their wide distribution. However, cultivated peanut is allotetraploid, with one gene theoretically having at least two copies (one from A genome and another from B genome), and some genes belong to a family of similar genes which adds to the genomic complexity. Thus, having more sequences available from various cultivated peanut species and tissues will be fundamental for new marker development and advancement of future peanut genomic research. Due to the paucity of peanut information in global repositories such as NCBI, only half of the pod abundant transcripts have been meaningfully annotated (Payton et al. 2009), and a large proportion of genes from root could not be functionally annotated. In most cases, we were unable to determine the exact function of a given molecule and could not annotate most of the *A. hypogaea* genes due to the fragmented nature of the EST clusters and large evolutionary differences from the model plant. Therefore, the underlying mechanism of resistance based on differential gene expression may be misleading. However, it is highly conceivable that, together with the acquisition and functional analysis of their full-length cDNA sequences rapidly accumulating in the public databases, we will be able to functionally characterize most of these genes in the near future.

The analysis of abundantly expressed genes in tissues indicated that there are major categories of genes expressed in specific tissues. In the developing peanut seeds, the decrease of carbohydrates has been shown to be

accompanied by the accumulation of crude fat, while the protein content does not change greatly (Pickett 1950). Our EST data from developing seed showed similar results. The majority of abundant transcripts in peanut seeds encoded storage proteins such as arachin and conarachin. The oleosin protein was also abundantly expressed. Oleosin plays an important role in the formation and stability of oil bodies (Chuang et al. 1996) and thus is of vital importance for oil accumulation in peanut seeds. At the same time, granule-bound starch synthase was also abundantly expressed in peanut seeds. Granule-bound starch synthase is responsible for the synthesis of amylose in starch granules (Denyer et al. 2001). These results indicated that protein and lipid formation in peanut is partially regulated at the transcriptional level, consistent with results of previously reported microarray analysis (Payton et al. 2009). In leaves, the abundantly expressed genes were predominantly related to photosynthesis, while auxin-repressed protein and metallothionein were also highly expressed. The findings were similar to those from leaves of grape, ginseng and oilseed rape (Guo et al. 2009; Lee et al. 1998). It was expected that the photosynthesis-related proteins were the most prevalent in leaves since they would facilitate the production of energy and materials for many activities such as cell division, growth, elongation, development and differentiation. Meanwhile, in peanut roots, molecules such as aquaporin, ABA-responsive protein, lectin, lipid transfer protein, catalase, glutathione *S*-transferase, PR10 and Ara h8 allergen were abundantly expressed. These types of genes are also upregulated in peanut pod subjected to fungal infection and drought stress (Luo et al. 2005c), and their expression may help the plant to tolerate the biotic and abiotic stresses frequently occurring in soil.

Altering fatty acid composition and increasing oil content are two major goals in plant lipid genetic engineering, as ~50 % of peanuts are crushed for oil production in China. Many potentially useful functional genes from peanut were isolated in this work such as *Fad 2*, *DGAT* and *oleosin*. Regulatory genes for oil content such as *lec1*, *fus3*, *ABI3*, *fus6*, *glabra2* were also isolated from peanut. The functional analysis of these genes in other plants such as *Arabidopsis* and *Brassica napus* (Wang et al. 2007) indicated that they play an important role in the regulatory process of oil formation, which may help us to understand the genetic background underlying the oil content trait and to manipulate their expression in order to improve peanut oil content. It should be noted that some genes are also biochemically important for fatty acid synthesis but not for oil formation, and thus have other uses such as in abiotic and biotic stress resistance (Upchurch 2008). The tissue-specific expression of some genes will shed light on the molecular function of these molecules. Some peanut

protein components have also aroused much health concerns because of allergenicity. In this study, one of the allergen genes, *Ara H8*, was the most abundantly expressed ESTs in peanut seeds. *Ara H8* has been described relatively recently, and it shows sequence homology to several pathogenesis-related (PR) proteins and may itself be a PR protein. This gene seemed to be expressed not only in seeds, but also in roots and leaves. When the leaves were challenged with *R. solanacearum*, *Ara H8* expression was greatly enhanced, which was quite different from other allergens genes expressed in developing seeds. This result was also slightly different from the result obtained in a previous study (Proite et al. 2007), which reported that it is expressed abundantly in roots of *A. stenosperma*. These results indicated that some peanut allergen genes may play a role in pathogenic resistance.

Activation of defense responses towards pathogen infection is associated with fine-tuned transcriptional regulation and the global alteration of gene expression patterns (Eulgem and Somssich 2007). Therefore, in this study, the gene expression pattern after *R. solanacearum* infection was analyzed to elucidate the possible mechanism of disease resistance. As the expression of many genes involved in the ethylene and JA signal transduction pathway were upregulated, the results suggest that they may be involved in the development of bacterial wilt resistance (Tables 4, 5). It has been reported previously that treatment with other microorganisms such as *Pythium oligandrum* can induce both the JA and ethylene signal transduction pathways, which may serve as resistance mechanisms against bacterial wilt (Hase et al. 2006, 2008). The involvement of ethylene and JA-mediated defense pathways in bacterial wilt resistance has also been suggested via virus-induced gene silencing of candidate genes (Chen et al. 2009). JA and ethylene have been demonstrated to promote the production of secondary metabolites in other plants (Dittrich et al. 1992). In this work, the expression of genes involved in secondary metabolism significantly increased after pathogenic attack (Table 6), which may further support the conclusion that the JA and ethylene signal transduction pathways participate in disease resistance. By contrast, based on the gene expression data in this work, it seems that the SA signal transduction pathway plays a small role in the resistance mechanism as the related genes did not change significantly after pathogenic attack. In contrast, in tomato, the exogenous chemical application to initiate SA signal transduction (Pradhanang et al. 2005) and the comparative analysis between resistant and susceptible lines that indicated SA plays an important role in the host plant mechanism of resistance (Afroz et al. 2009). These conflicting results further support the hypothesis that the resistance mechanism to *R. solanacearum* is diverse (Lin et al. 2008; Robatzek et al. 2007), and the genetic

improvement strategy to this disease should be host-specific. Therefore, even though resistance genes have been isolated from *Arabidopsis* (Deslandes et al. 2002), they cannot be used as candidate genes to enhance disease resistance in other crops. The analysis of gene expression in this study will contribute to the understanding of resistance mechanisms of peanut to bacterial wilt disease. The availability of this large set of sequences will also help to find the markers closely linked to bacterial wilt and eventually resistance genes.

Although many experimental studies are still needed to be performed in order to confirm the assumptions based on analysis of the expressed sequence data, this work can greatly contribute to (1) better understanding of the complexity of the peanut genetic background, (2) development of new markers such as SSR and SNP, (3) the isolation of candidate genes such as those with functions in fatty acid biosynthesis, TAG formation and bacterial wilt resistance, as well as regulatory and tissue-specific genes.

Conclusions

In the present study, 5 cDNA libraries with about 46 % clones containing full length genes have been constructed. A total of 63,234 peanut ESTs from peanut leaf, root and developing seeds have been sequenced. Transcription factor genes, tissue-specific genes, fatty acid biosynthesis and oil formation regulation genes, and RGA genes, as well as genes responsive to *R. solanacearum* infection have been obtained. These data will be served as valuable resources for peanut genomic work in the future.

Acknowledgments This work was supported by grants from the National Basic Research Program of China (2011CB109304-4), the National High Technology Research and Development Program of China (2006AA10A115) and Chinese Agricultural Research System (CARS-14).

References

- Afroz A, Khan MR, Ahsan N, Komatsu S (2009) Comparative proteomic analysis of bacterial wilt susceptible and resistant tomato cultivars. *Peptides* 30:1600–1607
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Asamizu E, Nakamura Y, Sato S, Tabata S (2000) Generation of 7137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. *DNA Res* 7:127–130
- Bi YP, Liu W, Xia H, Su L, Zhao CZ, Wan SB, Wang XJ (2010) EST sequencing and gene expression profiling of cultivated peanut (*Arachis hypogaea* L.). *Genome* 53:832–839
- Boote KJ (1982) Growth stages of peanut (*A. hypogaea* L.). *Peanut Sci* 9:35–40

- Chen YY, Lin YM, Chao TC, Wang JF, Liu AC, Ho FI, Cheng CP (2009) Virus-induced gene silencing reveals the involvement of ethylene-, salicylic acid- and mitogen-activated protein kinase-related defense pathways in the resistance of tomato to bacterial wilt. *Physiol Plant* 136:324–335
- Chi XY, Chen MC, Yang QL, He YN, Pan LJ, Gao Y, Yun SL (2010) Isolation and expression analysis of a β -ketoacyl-acyl carrier protein synthase I gene from *Arachis hypogaea* L. *Legume Genomics Genet* 1. doi:10.5376/lgg.2010.5301.0003
- Chu Y, Holbrook CC, Ozias-Akins P (2009) Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. *Crop Sci* 49:2029–2036
- Chuang RLC, Chen JCF, Chu J, Tzen JTC (1996) Characterization of seed oil bodies and their surface oleosin isoforms from rice embryos. *J Biochem* 120:74–81
- Dang PM, Guo B, Scully BT (2008) Identification and analysis of viral sequences in peanut (*Arachis hypogaea* L.) expressed sequence tags. *Phytopathology* 98:S44–S44
- Denyer K, Johnson P, Zeeman S, Smith AM (2001) The control of amylose synthesis. *J Plant Physiol* 158:479–487
- Deslandes L, Olivier J, Theulieres F, Hirsch J, Feng DX, Bittner-Eddy P, Beynon J, Marco Y (2002) Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. *Proc Natl Acad Sci USA* 99:2404–2409
- Dittrich H, Kutchan TM, Zenk MH (1992) The jasmonate precursor, 12-oxo-phytodienoic acid, induces phytoalexin synthesis in *Petroselinum crispum* cell cultures. *FEBS Lett* 309:33–36
- Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. *Curr Opin Plant Biol* 10:366–371
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185
- Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res* 9:950–959
- Fernandes J, Brendel V, Gai XW, Lal S, Chandler VL, Elumalai P, Galbraith DW, Pierson EA, Walbot V (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiol* 128:896–910
- Green P (2004) CrossMatch. <http://phrap.org/phrap.docs/general.html>
- Gueguen Y, Cadoret JP, Flament D, Barreau-Roumiguere C, Girardot AL, Garnier J, Hoareau A, Bachere E, Escoubas JM (2003) Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. *Gene* 303:139–145
- Guo BZ, Chen XP, Dang P, Scully BT, Liang XQ, Holbrook CC, Yu JJ, Culbreath AK (2008) Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection. *BMC Dev Biol* 8:12
- Guo BZ, Chen XP, Hong YB, Liang XQ, Dang P, Brennen T, Holbrook CC, Culbreath A (2009) Analysis of gene expression profiles in leaf tissues of cultivated peanuts and development of EST-SSR markers and gene discovery. *Int J Plant Genomics*. doi:10.1155/2009/715605
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258–D261
- Hase S, Shimizu A, Nakaho K, Takenaka S, Takahashi H (2006) Induction of transient ethylene and reduction in severity of tomato bacterial wilt by *Pythium oligandrum*. *Plant Pathol* 55:537–543
- Hase S, Takahashi S, Takenaka S, Nakaho K, Arie T, Seo S, Ohashi Y, Takahashi H (2008) Involvement of jasmonic acid signalling in bacterial wilt disease resistance induced by biocontrol agent *Pythium oligandrum* in tomato. *Plant Pathol* 57:870–876
- Jain S, Srivastava S, Sarin NB, Kav NNV (2006) Proteomics reveals elevated levels of PR 10 proteins in saline-tolerant peanut (*Arachis hypogaea*) calli. *Plant Physiol Biochem* 44:253–259
- Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34:D354–D357
- Kochert GD, Halward T, Stalker HT (1996) Genetic variation in peanut and its implications in plant breeding. In: Pickersgill B, Lock JM (eds) *Advances in legume science 8: legumes of economic importance*. Royal Botanic Gardens, Kew, pp 19–30
- Lee CM, Lee YJ, Lee MH, Nam HG, Cho TJ, Hahn TR, Cho MJ, Sohn U (1998) Large-scale analysis of expressed genes from the leaf of oilseed rape (*Brassica napus* L.). *Plant Cell Rep* 17:930–936
- Li HG, Wang L, Zhang YS, Lin XD, Liao B, Yan YS, Huang SZ (2005) Cloning and sequencing of the gene Ahy-beta encoding a subunit of peanut conarachin. *Plant Sci* 168:1387–1392
- Li MJ, Li AQ, Xia H, Zhao CZ, Li CS, Wan SB, Bi YP, Wang XJ (2009) Cloning and sequence analysis of putative type II fatty acid synthase genes from *Arachis hypogaea* L. *J Biosci* 34:227–238
- Li MJ, Wang XJ, Su L, Bi YP, Wan SB (2010a) Characterization of five putative acyl carrier protein (ACP) isoforms from developing seeds of *Arachis hypogaea* L. *Plant Mol Biol Rep* 28:365–372
- Li MJ, Xia H, Zhao CZ, Li AQ, Li CS, Bi YP, Wan SB, Wang XJ (2010b) Isolation and characterization of putative acetyl-CoA carboxylases in *Arachis hypogaea* L. *Plant Mol Biol Rep* 28:58–68
- Liang XQ, Chen XP, Hong YB, Liu HY, Zhou GY, Li SX, Guo BZ (2009) Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol* 9:35
- Liao BS (2005) A broad review and perspective on breeding for resistance to bacterial wilt. In: Allen C, Prior P, Hayward AC (eds) *Bacterial wilt disease and the Ralstonia solanacearum species complex*. The American Phytopathological Society, pp 225–238
- Lin YM, Chou IC, Wang JF, Ho FI, Chu YJ, Huang PC, Lu DK, Shen HL, Elbaz M, Huang SM, Cheng CP (2008) Transposon mutagenesis reveals differential pathogenesis of *Ralstonia solanacearum* on tomato and *Arabidopsis*. *Mol Plant Microbe Interact* 21:1261–1270
- Luo M, Dang P, Bausher MG, Holbrook CC, Lee RD, Lynch RE, Guo BZ (2005a) Identification of transcripts involved in resistance responses to leaf spot disease caused by *Cercosporidium personatum* in peanut (*Arachis hypogaea*). *Phytopathology* 95:381–387
- Luo M, Dang P, Guo BZ, He G, Holbrook CC, Bausher MG, Lee RD (2005b) Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut. *Crop Sci* 45:346–353
- Luo M, Liang XQ, Dang P, Holbrook CC, Bausher MG, Lee RD, Guo BZ (2005c) Microarray-based screening of differentially

- expressed genes in peanut in response to *Aspergillus parasiticus* infection and drought stress. *Plant Sci* 169:695–703
- Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, Bucher P, Copley RR, Courcelle E, Das U, Durbin R, Falquet L, Fleischmann W, Griffiths-Jones S, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lopez R, Letunic I, Lonsdale D, Silventoinen V, Orchard SE, Pagni M, Peyruc D, Ponting CP, Selengut JD, Servant F, Sigrist CJ, Vaughan R, Zdobnov EM (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res* 31:315–318
- Nobile PM, Lopes CR, Barsalobres-Cavallari C, Quecim V, Coutinho LL, Hoshino AA, Gimenes MA (2008) Peanut genes identified during initial phase of *Cercosporidium personatum* infection. *Plant Sci* 174:78–87
- Payton P, Kottapalli KR, Rowland D, Faircloth W, Guo BZ, Burow M, Puppala N, Gallo M (2009) Gene expression profiling in peanut using high density oligonucleotide microarrays. *BMC Genomics* 10:265
- Pickett TA (1950) Composition of developing peanut seed. *Plant Physiol* 25:210–214
- Pradhanang PM, Ji P, Momol MT, Olson SM, Mayfield JL, Jones JB (2005) Application of acibenzolar-S-methyl enhances host resistance in tomato against *Ralstonia solanacearum*. *Plant Dis* 89:989–993
- Proite K, Leal-Bertioli S, Bertioli D, Moretzsohn M, da Silva F, Martins N, Guimaraes P (2007) ESTs from a wild *Arachis* species for gene discovery and marker development. *BMC Plant Biol* 7:7
- Robatzek S, Bittel P, Chinchilla D, Kochner P, Felix G, Shiu SH, Boller T (2007) Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. *Plant Mol Biol* 64:539–547
- Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JC, Kerlavage AR (1996) The construction of Arabidopsis expressed sequence tag assemblies—a new resource to facilitate gene identification. *Plant Physiol* 112:1177–1183
- Saha S, Enugutti B, Rajakumari S, Rajasekharan R (2006) Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiol* 141:1533–1543
- Shoemaker R, Keim P, Vodkin L, Retzel E, Clifton SW, Waterston R, Smoller D, Coryell V, Khanna A, Erpelding J, Gai XW, Brendel V, Raph-Schmidt C, Shoop EG, Vielweber CJ, Schmatz M, Pape D, Bowers Y, Theising B, Martin J, Dante M, Wylie T, Granger C (2002) A compilation of soybean ESTs: generation and analysis. *Genome* 45:329–338
- Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B, Lundeberg J (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. *Proc Nat Acad Sci USA* 95:13330–13335
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637
- Tatusov R, Fedorova N, Jackson J, Jacobs A, Kiryutin B, Koonin E, Krylov D, Mazumder R, Mekhedov S, Nikolskaya A, Rao BS, Smirnov S, Sverdlov A, Vasudevan S, Wolf Y, Yin J, Natale D (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinforma* 4:41
- Tirumalaraju SV, Jain M, Gallo M (2011) Differential gene expression in roots of nematode-resistant and -susceptible peanut (*Arachis hypogaea*) cultivars in response to early stages of peanut root-knot nematode (*Meloidogyne arenaria*) parasitization. *J Plant Physiol* 168:481–492
- Upchurch R (2008) Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnol Lett* 30:967–977
- Wang HY, Guo JH, Lambert KN, Lin Y (2007) Developmental control of Arabidopsis seed oil biosynthesis. *Planta* 226:773–783
- Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. *Plant Mol Biol* 35:135–144
- Yan YS, Lin XD, Zhang YS, Wang L, Wu KQ, Huang SZ (2005) Isolation of peanut genes encoding arachins and conglutins by expressed sequence tags. *Plant Sci* 169:439–445
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:W293–W297
- Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848