Plant Physiol. (1998) 117: 1325–1332

Expressed Sequence Tags from a Root-Hair-Enriched *Medicago truncatula* cDNA Library[1]

Peter A. Covitz[2], Lucinda S. Smith, and Sharon R. Long*

Department of Biological Sciences (P.A.C., L.S.S., S.R.L.), and Howard Hughes Medical Institute (L.S.S., S.R.L.), Stanford University, Stanford, California 94305–5020

The root hair is a specialized cell type involved in water and nutrient uptake in plants. In legumes the root hair is also the primary site of recognition and infection by symbiotic nitrogen-fixing *Rhizobium* bacteria. We have studied the root hairs of *Medicago truncatula*, which is emerging as an increasingly important model legume for studies of symbiotic nodulation. However, only 27 genes from *M. truncatula* were represented in GenBank/EMBL as of October, 1997. We report here the construction of a root-hair-enriched cDNA library and single-pass sequencing of randomly selected clones. Expressed sequence tags (899 total, 603 of which have homology to known genes) were generated and made available on the Internet. We believe that the database and the associated DNA materials will provide a useful resource to the community of scientists studying the biology of roots, root tips, root hairs, and nodulation.

Roots provide structural and physiological support for plant interactions with the soil environment. Epidermal root hairs are specialized cells with a high surface-to-volume ratio that enables them to perform an important role in transport of water, ions, and nutrients. Root hairs are outgrowths of trichoblasts, and elongate by tip growth, a distinct mode of plant cell growth shared only by pollen tubes (Peterson and Farquhar, 1996). The patterning, differentiation, and growth of root hairs has been elucidated by genetic and cell biological studies (Di Cristina et al., 1996; Galway et al., 1997; Sanchez-Fernandez et al., 1997; Schneider et al., 1997), but many aspects of root-hair cell function remain unknown. In the Fabaceae, root hairs are also significant in having active responses to *Rhizobium* and related bacteria during the early stages of the symbiosis that leads to nitrogen fixation (Brewin, 1991; Hirsch, 1992). The identification of root proteins important for transport, cell growth and differentiation, and interaction with microbes is therefore of interest for studies in numerous plant species. However, because root hairs are small and often transient structures, direct biochemical analysis is challenging.

Medicago truncatula has emerged as an important experimental plant species both for studying nodulation by *Rhizobium* and for investigating mycorrhizal associations. *M. truncatula* is nodulated by *R. meliloti,* a bacterial species well characterized with respect to genetics and biochemistry. *M. truncatula* is autogamous and has a relatively small diploid genome and a number of genetically distinguishable ecotypes. These properties contribute to its suitability for molecular genetic analyses (Barker et al., 1990; Blondon et al., 1994; Cook et al., 1997). Genetic screens for plants with altered symbiotic phenotypes and various molecular approaches have identified several interesting mutations and genes (Benaben et al., 1995; Cook et al., 1995; Sagan et al., 1995; Gamas et al., 1996; Harrison, 1996; Peng et al., 1996; Burleigh and Harrison, 1997; Penmetsa and Cook, 1997).

Single-pass sequencing of cDNAs randomly picked from a library of genes made from a tissue of interest offers a complementary approach to biochemical and genetic analysis (Adams et al., 1991). ESTs generated by such an effort are compared with databases of identified genes. The results of these comparisons are used as a guide to assign putative identifications to the cDNAs. Researchers may then search or browse through the putative identifications of the ESTs to determine which genes may be of interest for further study. An investigator can also compare an experimentally obtained partial protein or nucleic acid sequence with the EST database to find out whether the cDNA for the gene has already been cloned. These methods have led to the rapid identification of genes in a number of organisms and have accelerated research by providing genetic material for further investigation (Adams et al., 1991, 1993; McCombie et al., 1992; Okubo et al., 1992; Newman et al., 1994; Rounsley et al., 1996).

In this paper we describe the collection of ESTs from a root-hair-enriched root-tip cDNA library from *M. truncatula*. We have constructed a website for access to the resulting database of 899 sequences. The sequence information and clones resulting from this effort may provide a useful tool for researchers studying general root physiology and molecular biology, and should have particular applications to the molecular biology of the *R. meliloti-M. truncatula* symbiosis.

[1] S.R.L. is an investigator of the Howard Hughes Medical Institute. Additional support for this project came from the Department of Energy, Energy Biosciences Program (contract no. DE-FG03-90ER20010). P.A.C. was a fellow of the Jane Coffin Childs Memorial Fund for Medical Research.

[2] Present address: Incyte Pharmaceuticals, 3174 Porter Drive, Palo Alto, CA 94304.

* Corresponding author; e-mail fa.srl@forsythe.stanford.edu; fax 1–650–725–8309.

Abbreviations: EST, expressed sequence tag.

MATERIALS AND METHODS

Tissue Collection

Sprouted seeds of *Medicago truncatula* cv Jemalong (Purkiss Seeds, Armidale, Australia) were grown overnight on 25 × 25-cm agar plates of Nod3 medium (2 mM $CaSO_4$, 1 mM $MgSO_4$, 0.5 mM K_2HPO_4, 1 mM 2-[*N*-morpholino]ethane sulfonic acid, pH 6.5, Murashige and Skoog minor salts without KI, and 11.5 g/L purified agar [Sigma] plus 1 μM Ag_2SO_4). Growth in enclosed plates was necessary to maintain sterility. However, under such conditions *M. truncatula* seedlings are sensitive to accumulated ethylene, an inhibitor of nodulation. Adding Ag_2SO_4, an inhibitor of plant responses to ethylene, to the growth medium blocked the effect of ethylene on nodulation (P.A. Covitz, unpublished results), and presumably reduced the likelihood that nodulation-related genes were improperly expressed due to ethylene.

Seedlings were placed on a sterile, wet paper towel with roots dangling 2 to 3 cm over one long edge, and the towel was rolled. The exposed root tips were dipped in liquid nitrogen and broken off. In one preparation these root tips with intact root hairs were used directly for RNA isolation. In another preparation root hairs were isolated by stirring in liquid nitrogen on a magnetic stir plate for 15 to 30 min until they broke off. Stripped root tips fell to the bottom of the container, whereas detached root hairs floated and were collected by decanting the liquid nitrogen (Rohm and Werner, 1987).

RNA Isolation

Tissue was ground under liquid nitrogen using a mortar and pestle and transferred to a centrifuge tube containing 8 mL of hot (70°C) borate extraction buffer (0.2 M sodium borate, 1% SDS, and 30 mM EGTA) per gram of tissue. An equal volume of Tris-EDTA-saturated phenol:chloroform (1:1, v/v), pH 9.0, was added. The mixture was vortexed for 1 min and put on ice for 5 min before homogenization with a polytron (model Kinematica PT 1200, Brinkmann Instruments, Westbury, NY). The sample was centrifuged and the aqueous layer was re-extracted three times with phenol:chloroform (1:1, v/v) and once with chloroform only. RNA was isolated by differential precipitation in 2 M LiCl followed by reprecipitation with ethanol.

cDNA Library Construction

Total RNA from root hairs (287 μg) and intact root tips with root hairs (663 μg) was pooled and sent to Stratagene for poly(A^+) RNA selection and construction of a root-hair-enriched root tip cDNA library. First-strand cDNA synthesis used an oligo-dT linker-primer with a *Xho*I cloning site. The 5′ end of each cDNA was ligated to an adaptor with an *Eco*RI-compatible overhang. cDNA was ligated unidirectionally into the *Eco*RI and *Xho*I sites of the λ-ZAP Express vector (Stratagene), packaged in vitro, and amplified. The amplified library represents approximately 10^6 recombinants.

Sequencing

The phage library was converted to the plasmid form by mass excision according to the procedure described by Stratagene. Amplified library lambda phage were co-infected with M13-derived ExAssist helper phage into *Escherichia coli* strain XL1-Blue MRF′, and the bacteria were grown for 2.5 h. The culture supernatant containing single-stranded phagemid form of the library was used to infect *E. coli* strain XLOLR. The bacteria were grown for 75 min and then used directly for double-stranded plasmid DNA preparation. Plasmid library DNA was electroporated into *E. coli* strain XL1-Blue, and the bacteria were plated at low density on medium containing Luria-Bertani broth, tetracycline (10 mg L^{-1}), and kanamycin (25 mg L^{-1}) after an outgrowth time of 40 min. Individual colonies were selected randomly for plasmid DNA purification and sequencing.

Template purification and sequencing of the plasmid cDNAs was performed either by the authors or by commercial DNA sequencing service providers (e.g. Bio-101 [Vista, CA] and Lark Technologies [Houston, TX]). All sequencing reactions contained the standard T3 sequencing primer, and thus read into the presumed 5′ end of each cDNA. Reactions were run and analyzed on either capillary or slab-gel electrophoresis automated sequencing machines (Perkin-Elmer). These machines generate two computer files for each sequencing run: a chromatogram file and a plain text file.

Sequence Editing

All stages of data analysis and assembly were performed on Macintosh operating system-based computers. The sequence text files were edited to remove leading vector and trailing, poor-quality sequence using the Java-based computer program SeqTrim, which was written for this work (P.A. Covitz, unpublished program). SeqTrim also flagged anomalous clone sequences that were then edited manually after examination of their corresponding chromatogram files.

Homology Comparisons and Database Construction

Edited EST sequences were entered into FileMaker Pro (Claris, Santa Clara, CA), a relational database. Each EST was translated in all six reading frames and compared with the nonredundant database at the National Center for Biotechnology Information (NCBI) using the BLASTX program. Default BLAST parameter values were used except for the following settings: Expect = 1, Alignments = 3, and Descriptions = 10. Sequences that returned no significant homology were again compared using BLASTN with Expect = 0.1, Alignments = 3, and Descriptions = 10. The results of the comparisons were incorporated into the FileMaker database. Homologies to negative reading frames were disregarded, except in clones with inserts in the reverse orientation. Putative identifications for the ESTs were assigned based on the results of the BLAST searches and in some cases with information contained in related abstracts in MEDLINE. WebStar (Quarterdeck) and Tango for File-

Maker (Everyware) are being used to display the FileMaker database to users on the World Wide Web.

RESULTS

Single-Pass Sequencing of Random cDNAs

Root-hair cells are the site of infection by *R. meliloti*, yet they represent only a small proportion of the total mass of tissue in the root. We postulated that genes uniquely or preferentially expressed in root hairs may be critical for symbiotic recognition. We therefore constructed a cDNA library from root tissue enriched for root hairs to increase the proportionate representation of root-hair-specific genes. The RNA source material for the library was derived primarily from the infection zone of the root and contained growing root hairs, fully differentiated root hairs, root tips including the meristem, and root-cap cells (see "Materials and Methods"). Thus, all of the major cell types and processes related to the early stages of symbiosis were represented.

We constructed the library by unidirectional insertion of oligo-dT-primed cDNAs into λ-ZAP-Express. The phage library was converted through mass excision to a plasmid library in the vector pBK-CMV. Individual clones from the plasmid library were picked at random and sequenced using a standard T3 primer that was annealed and extended into the 5′ side of the inserted cDNA. Of 958 sequencing reactions attempted, 919 produced some length of readable sequence.

The text file resulting from each sequencing reaction was edited using both automated and, if necessary, manual methods. SeqTrim was used to remove leading vector and trailing, poor-quality sequences from each file. In cases in which the sequencing reaction extended past the poly(A^+) tail of the gene, the sequence was trimmed to remove the 3′ vector and linker sequences. SeqTrim also flagged sequences derived from anomalous clones. These anomalies generally fell into three classes: (a) clones with altered adaptor sequence or incorrect base calls at the junction between the vector and the cDNA insert; (b) clones with inserts in the reverse orientation; or (c) clones without inserts altogether. The files for these three classes were edited manually; sequences from the third class were removed from further consideration.

Putative Identification of Genes

Each EST was compared against all sequences in the nonredundant database at the NCBI using the program BLASTX, which compares translated nucleotide sequences with protein sequences. Sequences that had no homology to any protein in the database were then reanalyzed using the program BLASTN, which compares nucleotide sequences with nucleotide sequences. The results of each comparison were screened manually. Sequences deemed to be of bacterial origin were removed from the collection. After screening and editing we retained 899 ESTs. Statistical information about the collection is shown in Table I.

Table I. *EST collection statistics*

EST length (nucleotides)	
Mean	601
Median	604
SD	71
Range	221–720
Ambiguous base calls (percent per EST)	
Mean	1.6
Median	0.9
cDNA end sequenced (no. of ESTs)	
5′	889
3′	10
Homology comparisons (no. of ESTs)	
Similar to known genes	603
Novel	296

Of the 899 ESTs, 603 had significant homology to previously identified genes. Although the BLAST scores and P values were considered, the assessment of whether a given homology was significant was determined by investigator judgment, not by absolute numerical cutoffs. The annotations of genes with similarities to an EST were used to assign a putative identification to that EST. In cases in which the annotation was vague, information contained in MEDLINE abstracts related to the gene was used to assign a putative identification. The 603 ESTs with similarity to known genes represent 356 distinct genes, as indicated by their putative identifications. These distinct identifications were grouped into 13 functional categories, which are listed in Table II.

Abundantly Expressed Genes

The relative abundance of the mRNA in a tissue is approximately reflected in the abundance of its corresponding cDNA in non-normalized libraries. Random sequencing of cDNAs therefore yields information about the relative expression levels of the genes represented by the ESTs (Adams et al., 1993). Table III lists the genes from the five most abundant mRNAs in our library. Two of these, Met synthase and elongation factor 1-61, are critical for protein metabolism, whereas β-glucosidase is involved in cell wall development. The abundant expression of these genes reflects the actively growing state of the source tissue used to generate the library. The remaining two most abundantly expressed genes are members of the membrane intrinsic protein water channel family, a result that is consistent with the physiological role of the root in water uptake.

Met synthase is also related to ethylene metabolism, so the abundance of the gene is intriguing because of the possibly complex role(s) of ethylene in root-hair development and symbiosis (Masucci and Schiefelbein, 1996; Dolan, 1997; Heidstra et al., 1997; Penmetsa and Cook, 1997). However, the presence of the ethylene-response inhibitor Ag_2SO_4 in the plant growth medium (see "Materials and Methods") may have perturbed the expression of enzymes involved in ethylene metabolism. This possibility was not specifically tested in *M. truncatula* roots.

Table II. *EST putative identifications*

- Cell wall structure or metabolism
  - Arabinogalactan protein
  - β-Galactosidase
  - β-Glucosidase
  - Cellulase
  - Cellulose synthase
  - Extensin
  - Extracellular dermal glycoprotein
  - Hyp-rich protein
  - Pectin acetylesterase
  - Pectin methylesterase
  - Pollen-specific protein
  - Polygalacturonase
  - Pro-rich protein
- Cytoskeleton
  - Actin
  - Actin depolymerizing factor
  - Ankyrin-repeat protein
  - Annexin
  - F-Actin capping protein α-subunit
  - Myosin
  - Profilin
  - Tubulin α chain
  - Tubulin β chain
- Membrane transport
  - ADP, ATP carrier protein
  - ATP-dependent transporter
  - Glc transporter
  - Membrane channel protein
  - Mitochondrial phosphate transporter
  - *Myo*-inositol transporter
  - Plasma membrane intrinsic protein
  - Sugar transporter
  - Tonoplast intrinsic protein
  - Tonoplast intrinsic protein Δ
  - Tonoplast intrinsic protein γ
  - Transporter
  - Triosephosphate translocator
  - Vacuolar H^+ ATPase subunit A
  - Vacuolar H^+ ATPase subunit B
  - Water-stress-induced tonoplast intrinsic protein
- Signal transduction
  - 14–3–3 Protein
  - 2A6,1-Aminocyclopropane-1-carboxylate oxidase related
  - Adenylyl cyclase-associated protein 2
  - ADP-ribosylation factor; ARF
  - Calcium-dependent protein kinase
  - Calmodulin
  - Calreticulin
  - Casein kinase II
  - Eyes absent
  - Fimbriata
  - GTP-binding protein
  - HASPP28, casein kinase II substrate
  - Int-6 tumor-related protein
  - Leu-rich repeat receptor protein kinase
  - Light-repressible receptor protein kinase
  - Nedd1
  - Neuronal calcium sensor 1
  - Nucleotide phosphatase; apyrase
  - Phosphatidylinositol 3-kinase
  - Phytochrome
  - Protein kinase
  - Protein kinase Wpk4, cytokinin and nutrient regulated
  - Protein phosphatase 2A regulatory subunit A
  - Protein phosphatase 2A regulatory subunit B
  - Ran-binding protein
  - Ran GTP-binding protein
  - Receptor protein kinase
  - Receptor protein kinase, S-locus related
  - Retinoblastoma-binding protein p46
  - Retinoblastoma-binding protein p48
  - Sensory transduction His kinase
  - Switch-activating protein related
  - Trimeric G protein β-subunit
- Cell division cycle
  - Cell cycle Ser/Thr protein kinase
  - CDC21
  - CDC48
  - CDC68
  - Cyclin C
  - Regulator of chromosome condensation 1
  - SKP1, kinetochore-binding cell cycle regulator
- Defense
  - Endochitinase
  - HSR203J plant defense gene
  - Hypersensitive response-inducing protein
  - Immediate-early salicylate-induced glucosyltransferase
  - Peroxidase
  - TMV resistance protein N
- Vesicular trafficking, protein sorting, and secretion
  - ER retention receptor
  - Importin
  - Low-density lipoprotein C; brefeldin A-sensitive Golgi protein
  - Nuclear localization signal-binding protein
  - Nuclear pore-targeting protein, 97-kD subunit
  - Rab G protein
  - Rab7G protein
  - Rab11G protein
  - Sec7 protein transport protein
  - Sec61 protein transport protein
  - Suppressor of clathrin deficiency 6
  - Translocon-associated protein, α-subunit
  - Vacuolar protein-sorting protein
  - Vacuolar sorting receptor
- Chromatin and DNA metabolism
  - DNA damage repair/toleration Leu-rich repeat protein
  - DNA helicase
  - DNA topoisomerase II
  - High-mobility group DNA-binding protein
  - Histone H2A
  - Histone H2B
  - Histone H4
  - Nucleosome assembly protein 1
  - Photolyase
  - UV excision repair protein
- Gene expression and RNA metabolism
  - CCAAT box DNA-binding transcription factor
  - Fibrillarin
  - Fork head rRNA processing protein
  - General negative regulator of transcription
  - GLO3 zinc finger protein
  - Gly-rich RNA-binding protein
  - Heterogeneous nuclear ribonucleoprotein K
  - Homeobox protein
  - Nucleolin
  - Nucleolysin TIAR
  - Poly(A^+) RNA export protein
  - Poly(A^+)-binding protein
  - Ribonucleoprotein
  - RING zinc finger protein
  - RNA-binding protein
  - RNA helicase
  - SEB4 ribonucleoprotein
  - Sm; small nuclear ribonucleoprotein
  - Spliceosome-associated protein
  - Splicing factor
  - SSN6; transcriptional repressor
  - TATA-box binding protein TBP
  - Transcription factor HBP-1b
  - Transcription initiation factor TFIID subunit
  - U1 and U2 small nuclear ribonuclear protein E
- Secondary and hormone metabolism
  - 17-β-Hydroxysteroid dehydrogenase
  - 4-Coumarate:CoA ligase
  - Auxin-induced glutathione *S*-transferase
  - Caffeoyl-CoA *O*-methyltransferase
  - Chalcone synthase
  - Cyt P-450
  - Δ 24-Sterol-methyltransferase
  - Dihydroflavonol-4-reductase
  - Farnesyl pyrophosphate synthase
  - GA4; 3 β-hydroxylase in GA synthesis
  - Glutathione *S*-transferase
  - Indole-3-acetate β-glucosyltransferase
  - Lipoxygenase
  - Squalene monooxygenase
- Primary metabolism
  - 1-Acyl-glycerol-3-phosphate acyltransferase
  - 3-Oxoacyl-[acyl-carrier-protein] reductase
  - 3-Phosphoshikimate 1-carboxyvinyltransferase
  - 6-Phosphogluconate dehydrogenase
  - 8-Amino-7-oxononanoate synthase
  - Acid phosphatase
  - Aconitate hydratase
  - Adenosine kinase
  - Adenosylhomocysteinase
  - Adenylate kinase
  - ADP-Glc synthase
  - Ala aminotransferase
  - Alcohol dehydrogenase
  - Aldolase
  - Argininosuccinate synthase
  - Aromatic-L-amino acid decarboxylase
  - Ascorbate peroxidase
  - ATP citrate lyase
  - ATP synthase
  - ATP synthase β chain
  - β-Succinyl CoA synthetase
  - Biotin synthase
  - Citrate synthase

Table II. *Continued*

- cobW
- Copper amine oxidase
- CTP synthase
- Cys lyase
- Cys synthase
- Cyt *b5*
- Cyt *c* oxidase
- Dihydrofolate reductase-thymidylate synthase
- Dihydroneopterin aldolase
- dTDP-Glc 4,6-dehydratase
- Enolase
- Epoxide hydrolase
- Fatty acid elongase
- Fru-1,6-bisphosphatase
- Fru-1,6-bisphosphate aldolase
- γ Glutamyl hydrolase
- Glc-6-phosphate isomerase
- Glucosyltransferase
- Glyceraldehyde-3-phosphate dehydrogenase
- Glyoxysomal citrate synthase
- Hydroxymethylglutaryl-CoA synthase
- Hydroxymethyltransferase
- Isocitrate dehydrogenase
- Isopentenyl pyrophosphate isomerase
- Ketol-acid reductoisomerase
- L-3-Hydroxyacyl-CoA hydrolyase, L-3-hydroxyacyl-CoA dehydrogenase-D-3-hydroxyacyl-CoA epimerase, and Δ 3,Δ 2-enoyl-CoA isomerase-tetrafunctional protein
- L-Lactate dehydrogenase
- Leu aminopeptidase
- Lysophospholipase
- Malate dehydrogenase
- Met synthase
- Methylenetetrahydrofolate reductase
- Methyltransferase
- Mevalonate kinase
- NADH dehydrogenase
- NADH-Cyt *b5* reductase
- NFS1, pyridoxal phosphate-dependent aminotransferase
- *O*-methyltransferase
- Ornithine carbamoyltransferase
- Phenylalanine ammonia-lyase
- Phosphatidylinositol synthase
- PEP carboxykinase
- Phosphofructokinase
- Phosphoglucomutase
- Phosphoglycerate dehydrogenase
- Phosphoglycerate kinase
- Phosphoglycerate mutase
- Phosphomevalonate kinase
- Pro dehydrogenase
- Pro oxidase
- Pyrophosphatase
- Pyruvate dehydrogenase E1 α-subunit
- Pyruvate kinase
- Ribose 5-phosphate isomerase
- *S*-adenosylmethionine synthase
- Ser palmitoyltransferase
- Succinate dehydrogenase
- Suc synthase
- Thioredoxin
- Thymidine diphosphoglucose 4,6-dehydratase
- Transaldolase
- Transketolase
- Triosephosphate isomerase
- Tryptophan synthase α chain
- Ubiquinol-Cyt *c* reductase complex subunit VI-requiring protein
- UDP-Glc dehydrogenase
- Urease accessory protein G
- UTP-Glc glucosyltransferase
- Z-crystallin; NADPH:quinone reductase

Protein synthesis and processing

- 26S Protease regulatory subunit 8
- 26S Protease subunit Tat-binding protein 1
- 26S Protease subunit TBP-1; Mg-dependent ATPase
- 26S rRNA
- 40S Ribosomal protein S2
- 40S Ribosomal protein S3
- 40S Ribosomal protein S3A
- 40S Ribosomal protein S4
- 40S Ribosomal protein S5
- 40S Ribosomal protein S6
- 40S Ribosomal protein S7
- 40S Ribosomal protein S8
- 40S Ribosomal protein S9
- 40S Ribosomal protein S13
- 40S Ribosomal protein S14
- 40S Ribosomal protein S15
- 40S Ribosomal protein S17
- 40S Ribosomal protein S18
- 40S Ribosomal protein S19
- 40S Ribosomal protein S25
- 60S Acidic ribosomal protein P1
- 60S Acidic ribosomal protein P2
- 60S Ribosomal protein L1
- 60S Ribosomal protein L1A
- 60S Ribosomal protein L2
- 60S Ribosomal protein L3
- 60S Ribosomal protein L5
- 60S Ribosomal protein L6
- 60S Ribosomal protein L7
- 60S Ribosomal protein L7A
- 60S Ribosomal protein L9
- 60S Ribosomal protein L10A
- 60S Ribosomal protein L11
- 60S Ribosomal protein L12
- 60S Ribosomal protein L13
- 60S Ribosomal protein L13A
- 60S Ribosomal protein L17
- 60S Ribosomal protein L18A
- 60S Ribosomal protein L19
- 60S Ribosomal protein L24
- 60S Ribosomal protein L28
- 60S Ribosomal protein L29
- 60S Ribosomal protein L30
- 60S Ribosomal protein L31
- 60S Ribosomal protein L32
- Aminoacylase I
- Asparaginyl-tRNA synthetase
- BiP heat-shock protein
- Calnexin
- Chaperonin
- Chaperonin 60 β
- Cyclophilin
- Cys proteinase
- Deubiquitinating enzyme
- Endoplasmin; heat-shock protein 90
- Heat-shock cognate protein 70
- Heat-shock protein
- Heat-shock protein 70
- Heat-shock protein 90
- Kunitz-type protease inhibitor
- Leader peptidase I
- Lysyl-tRNA synthetase
- Methionyl-tRNA synthetase
- Mitochondrial chaperonin HSP60
- Mitochondrial processing peptidase
- Mitochondrial ribosomal protein
- Oligopeptidase A
- p40, ribosome associated; laminin receptor
- Peptidylprolyl isomerase; FK506-binding protein
- Proteasome subunit C2
- Proteasome subunit C3
- Proteasome subunit p112
- Protein disulfide isomerase
- PSTI-type protease inhibitor
- rof1, FK506-binding protein
- Ser carboxypeptidase
- Ser carboxypeptidase II
- Stress inducible; heat-shock protein
- Subtilisin-like protease
- Translation elongation factor 1-α
- Translation elongation factor 1-β
- Translation elongation factor 1-γ
- Translation elongation factor 2
- Translation elongation factor TU
- Translation initiation factor 2-α
- Translation initiation factor 4A
- Translation initiation factor 5A
- Ubiquitin
- Ubiquitin activating enzyme
- Ubiquitin conjugating enzyme
- Ubiquitin conjugating enzyme E2

Miscellaneous

- Apospory-associated protein
- AUX28 auxin-induced protein
- Chlorophyll *a*/*b*-binding protein
- DAG chloroplast differentiation protein
- Defective chloroplasts and leaves; DCL
- Diminuto, cell-elongation regulator
- Drought-induced 19-kD protein
- Early light-induced protein
- Embryo-abundant protein
- Embryogenesis-associated protein
- Gln repeat protein
- HvB12D
- IAA8; IAA9; auxin inducible
- KE2
- Late embryogenesis-abundant protein
- Leader open reading frame of *S*-adenosylmethionine decarboxylase gene
- LEC14B
- Leginsulin
- PSII type I chlorophyll *a*/*b*-binding protein
- Plant organ-specific protein
- Seed imbibition protein Sip1
- Translationally controlled tumor protein
- wali7 aluminum-induced gene

Table III. *Most abundant mRNAs*

Values in parentheses indicate percentage of total.

Putative Identification	No. of ESTs
Met synthase	12 (1.3)
β-Glucosidase	12 (1.3)
Water-stress-induced tonoplast intrinsic protein	9 (1.0)
Plasma membrane intrinsic protein	8 (0.9)
Translation elongation factor 1-α	7 (0.8)

DISCUSSION

Genes of Interest

Our RNA preparation included actively growing and differentiating cells. We therefore expected a diverse representation of sequences from cytoskeleton, vesicle trafficking, and cell division functions. Representatives of these classes were found, along with a number of sequences that encode likely cell wall proteins and cell wall synthesis enzymes. An even larger number of gene products with homology to proteins involved in signal transduction were identified. Considering the dynamic growth of root hairs and the induction of root cortical cell division in response to *R. meliloti* cells and Nod factors, the proteins encoded by sequences in each of these categories might be especially interesting targets for cell biological manipulation and analysis.

Genes categorized under defense and secondary metabolism functions should also be appropriate targets of study given the possible involvement, or suppression, of host defenses during infection by *Rhizobium* (Hirsch and Fang, 1994; Mellor and Collinge, 1995; Spaink, 1995). In particular, an endochitinase homolog may be intriguing to examine (EST 00194; serial nos. can be used to obtain detailed EST information at http://bio-SRL8.stanford.edu, as described below). This endochitinase EST does not have significant homology to a previously identified *M. truncatula* chitinase cDNA derived from roots infected with *R. meliloti* (accession no. Y10373; A. Niebel, unpublished data). The identification of two different chitinases in uninfected and infected roots permits more rigorous testing of the hypothesis that chitinases have a regulatory role in plant responses to lipooligosaccharide nodulation factors (Staehelin et al., 1994).

Several other sequences are especially interesting given their homology to genes with known functions in other systems. One intriguing sequence in the collection is a homolog of the mammalian *eyes absent* (*eya*) gene (EST 00777). The BLAST comparison of the EST with homology to this gene yielded high scoring pairs spanning 103 amino acids of human *eya2.* The P value of the match was $4.9e^{-14}$, suggesting that the alignment was not due to chance. The *eya* genes have been implicated in eye development in both mammals and insects, a finding that has challenged the long-held notion that eyes evolved independently in these two branches of animal phylogeny (Duncan et al., 1997; Zimmerman et al., 1997). The presence of an *eya* homolog in plants suggests that the history of this gene family may extend to a common ancestor of plants and animals.

Another sequence of particular significance is a His kinase (EST 00711) with the strongest similarity to a bacterial two-component regulator ($P < 1.1e^{-26}$). Only a few His kinase sequences have been reported in plants, but these are of high interest because one, the *ETR1* gene, has been identified by genetics as an ethylene receptor, and its product has been shown by direct biochemical studies to be an ethylene-binding protein (Chang et al., 1993; Schaller and Bleecker, 1995). Another sequence with His kinase homology is implicated by genetic tests as a possible cytokinin receptor (Kakimoto, 1996). Thus, the functionally characterized His kinase-like sequences reported in plants are possible receptors for chemical ligands. The appearance of a new His kinase in the root-hair-enriched library is important in light of the internal and external chemical signals that are operating in roots and their root hairs. The *M. truncalata* sequence appears to be distinct from *ETR1* and other functionally characterized genes. We are pursuing such characterization of this newly identified His kinase homolog through construction of antisense and other transgenic plants.

We expected to find expressed sequences representing major root functions such as transport and cell growth and division; however, since no cells from other portions of the plant were present, we did not expect to find functions typically associated with shoot, leaf, or reproductive organs. Therefore, some EST sequences in the library are developmentally surprising. These include chlorophyll *a*- and *b*-binding protein (ESTs 00122, 00414, and 00460), an embryo-specific protein (EST 00339), and a pollen-specific protein (EST 00010).

Several genes with homology to sugar transporters were found. No homologs of the nitrate transporter or of other putative plasmalemma transporters were identified. The apparent paucity of transporter genes is somewhat puzzling. However, we note that more than one-fifth of the ESTs in the database show no significant homology to known genes. Transporter genes may be represented among these unclassified sequences. It is also possible that the putative identifications of the membrane transporters in Table II reflect their basic membrane transport functions, but do not necessarily identify the correct substrate.

Sequence similarity as indicated by BLAST does not always reflect actual conservation at the relevant functional sites of the proteins in question. Examination of individual sequence lineups and reference to all original papers should precede conclusions about the likely function of any particular expressed gene. The Web site described below is designed to facilitate this process.

Internet Access to Detailed EST Information

The entire collection of ESTs has been organized into an online database that is accessible via the World Wide Web at http://bio-SRL8.stanford.edu. This Web site provides tools for browsing and searching the database. Each EST has a detailed record that includes the results and date of its BLAST comparison and links to additional information on the matching genes at the NCBI. This provides a way to examine the relationship of the putative homolog to the

gene being queried. The raw sequence chromatogram files and chromatogram-viewing software are available for downloading at this site as well. The raw data should prove useful to researchers who want to confirm the base calls of a particular sequence, for example, to design primers. In addition, all of the EST sequences have been deposited in dbEST at the NCBI. An investigator who wishes to compare his or her own sequence with these *M. truncatula* ESTs can do so by performing a BLAST search with the NCBI server by selecting dbEST as the database to search against (http://www.ncbi.nlm.nih.gov/BLAST).

Uses for the ESTs

The EST data from the *M. truncatula* cDNA library described here can initially be used to create codon-usage tables and other data tables to assist in the establishment of *M. truncatula* as a model system for molecular genetic studies. The sequences can be used to generate probes to isolate genomic DNA containing the corresponding genes and to provide markers for physical maps. Gene-expression studies may identify genes with cell-type-specific or symbiotically regulated expression patterns. Once isolated from genomic DNA, the promoters of such genes may provide valuable reagents for transgenic promoter-fusion experiments. Other genes described here may be useful as controls for constitutive expression.

Root hairs are mechanically and optically accessible and their growth can be actively studied using a number of cell biological techniques and experimental manipulations (Dazzo et al., 1996; Ehrhardt et al., 1996; Galway et al., 1997). EST sequences that encode proteins of known or predicted function may be used to create peptide antigens for generating antibodies to be used in such studies. The protein products of cytoskeletal protein and cell wall enzyme genes may be good candidates for this approach.

Finally, the EST database may be of use to scientists who have biochemically purified proteins of interest from *M. truncatula*. The partial peptide sequence of a purified protein could be compared against translated EST sequences. If present, related and more extensive cDNAs could then be readily identified and used as tools for additional studies.

ACKNOWLEDGMENTS

We are grateful to Audrey Southwick for assistance in the preparation of plants and isolation of RNA for the construction of the library and to Melanie Ukanwa for assisting with the BLAST homology searches. We thank Michael Cherry and David Flanders for their advice on constructing the EST database, JoAnne Connelly for assistance with the manuscript, and members of our laboratory for numerous useful suggestions.

Received December 30, 1997; accepted May 4, 1998.
Copyright Clearance Center: 0032–0889/98/117/1325/08.

LITERATURE CITED

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos M, Xiao H, Merril C, Wu A, Olde B, Moreno R, and others (1991) Complementary DNA sequencing expressed sequence tags and the human genome project. Science **252:** 1651–1656

Adams MD, Kerlavage AR, Fields C, Venter JC (1993) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nat Genet **4:** 256–267

Barker D, Bianchi S, Blondon F, Dattee Y, Duc G, Flament P, Gallusci P, Genier P, Guy P, Muel X, and others (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. Plant Mol Biol Rep **8:** 40–49

Benaben V, Duc G, Lefebvre V, Huguet T (1995) TE7, an inefficient symbiotic mutant of *Medicago truncatula* Gaertn. cv Jemalong. Plant Physiol **107:** 53–62

Blondon F, Marie D, Brown S, Kondorosi A (1994) Genome size and base composition in *Medicago sativa* and *M. truncatula* species. Genome **37:** 264–270

Brewin, N J (1991) Development of the legume root nodule. Annu Rev Cell Biol **7:** 191–226

Burleigh SH, Harrison MJ (1997) A novel gene whose expression in *Medicago truncatula* roots is suppressed in response to colonization by vesicular-arbuscular mycorrhizal (VAM) fungi and to phosphate nutrition. Plant Mol Biol **34:** 199–208

Chang C, Kwok SF, Bleecker AB, Meyerowitz EM (1993) Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. Science **262:** 539–544

Cook D, Dreyer D, Bonnet D, Howell M, Nony E, Vandenbosch K (1995) Transient induction of a peroxidase gene in *Medicago truncatula* precedes infection by *Rhizobium meliloti*. Plant Cell **7:** 43–55

Cook DR, Vandenbosch K, De Bruijn FJ, Huguet T (1997) Model legumes get the nod. Plant Cell **9:** 275–281

Dazzo FB, Orgambide GG, Philip-Hollingsworth S, Hollingsworth RI, Ninke KO, Salzwedel JL (1996) Modulation of development, growth dynamics, wall crystallinity, and infection sites in white clover root hairs by membrane chitolipooligosaccharides from *Rhizobium leguminosarum* biovar trifolii. J Bacteriol **178:** 3621–3627

Di Cristina M, Sessa G, Dolan L, Linstead P, Baima S, Ruberti I, Morelli G (1996) The Arabidopsis Athb-10 (GLABRA2) is an HD-zip protein required for regulation of root hair development. Plant J **10:** 393–402

Dolan, L (1997) The role of ethylene in the development of plant form. J Exp Bot **48:** 201–210

Duncan MK, Kos L, Jenkins NA, Gilbert DJ, Copeland NG, Tomarev (1997) Eyes absent: a gene family found in several metazoan phyla. Mammalian Genome **8:** 479–485

Ehrhardt DW, Wais R, Long SR (1996) Calcium spiking in plant root hairs responding to rhizobium nodulation signals. Cell **85:** 673–681

Galway ME, Heckman JW, Jr, Schiefelbein JW (1997) Growth and ultrastructure of Arabidopsis root hairs: the rhd3 mutation alters vacuole enlargement and tip growth. Planta **201:** 209–218

Gamas P, Niebel FDC, Lescure N, Cullimore JV (1996) Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol Plant Microbe Interact **9:** 233–242

Harrison MJ (1996) A sugar transporter from *Medicago truncatula*: altered expression pattern in roots during vesicular-arbuscular (VA) mycorrhizal associations. Plant J **9:** 491–503

Heidstra R, Yang WC, Yalcin Y, Peck S, Emons A, Van Kammen A, Bisseling T (1997) Ethylene provides positional information on cortical cell division but is not involved in Nod factor-induced root hair tip growth in rhizobium-legume interaction. Development **124:** 1781–1787

Hirsch A (1992) Tansley review no. 40: developmental biology of legume nodulation. New Phytol **122:** 211–237

Hirsch AM, Fang Y (1994) Plant hormones and nodulation: what's the connection? Plant Mol Biol **26:** 5–9

Kakimoto T (1996) CKI1, a histidine kinase homolog implicated in cytokinin signal transduction. Science **274:** 982–985

McCombie WR, Adams MD, Kelley JM, Fitzgerald MG, Utterback TR (1992) *Caenorhabditis elegans* expressed sequence tags

identify gene families and potential disease gene homologues. Nat Genet **1:** 124–131

Masucci JD, Schiefelbein JW (1996) Hormones act downstream of TTE and GL2 to promote root hair outgrowth during epidermis development in the Arabidopsis root. Plant Cell **8:** 1505–1517

Mellor RB, Collinge DB (1995) A simple model based on known plant defence reactions is sufficient to explain most aspects of nodulation. J Exp Bot **46:** 1–18

Newman T, De Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville C (1994) Genes galore. A summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones. Plant Physiol **106:** 1241–1255

Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nat Genet **2:** 173–179

Peng HM, Dreyer DA, Vandenbosch KA, Cook D (1996) Gene structure and differential regulation of the rhizobium-induced peroxidase gene rip1. Plant Physiol **112:** 1437–1446

Penmetsa RV, Cook DR (1997) A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. Science **275:** 527–530

Peterson, R L and Farquhar, M L (1996) Root hairs: specialized tubular cells extending root surfaces. Bot Rev **62:** 1–40

Rohm M, Werner D (1987) Isolation of root hairs from seedlings of *Pisum sativum*: identification of root hair specific proteins by in situ labeling. Physiol Plant **69:** 129–136

Rounsley SD, Glodek A, Sutton G, Adams MD, Somerville CR, Venter JC, Kerlavage AR (1996) The construction of Arabidopsis expressed sequence tag assemblies. Plant Physiol **112:** 1177–1183

Sagan M, Morandi D, Tarenghi E, Duc G (1995) Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after gamma-ray mutagenesis. Plant Sci **111:** 63–71

Sanchez-Fernandez R, Fricker M, Corben LB, White NS, Sheard N, Leaver CJ, Van Montagu M, Inze D, May MJ (1997) Cell proliferation and hair tip growth in the Arabidopsis root are under mechanistically different forms of redox control. Proc Natl Acad Sci USA **94:** 2745–2750

Schaller GE, Bleecker AB (1995) Ethylene-binding sites generated in yeast expressing the Arabidopsis ETR1 gene. Science **270:** 1809–1811

Schneider K, Wells B, Dolan L, Roberts K (1997) Structural and genetic analysis of epidermal cell differentiation in Arabidopsis primary roots. Development **124:** 1789–1798

Spaink HP (1995) The molecular basis of infection and nodulation by rhizobia: the ins and outs of sympathogenesis. Annu Rev Phytopathol **33:** 345–368

Staehelin C, Schultze M, Kondorosi E, Mellor RB, Boller T, Kondorosi A (1994) Structural modifications in *Rhizobium meliloti* Nod factors influence their stability against hydrolysis by root chitinases. Plant J **53:** 319–330

Zimmerman JE, Bui QT, Steingrimsson E, Nagle DL, Fu W, Genin A, Spinner NB, Copeland NG, Jenkins NA, Bucan M, and others (1997) Cloning and characterization of two vertebrate homologs of the *Drosophila* eyes absent gene. Genome Res **7:** 128–141