Construction of a *Lotus japonicus* Late Nodulin Expressed Sequence Tag Library and Identification of Novel Nodule-Specific Genes

Krzysztof Szczyglowski²*, Dirk Hamburger², Philipp Kapranov, and Frans J. de Bruijn

Department of Energy Plant Research Laboratory (K.S., D.H., P.K., F.J.d.B.), Genetics Program (P.K., F.J.d.B.), and Department of Microbiology (F.J.d.B.), Michigan State University, East Lansing, Michigan 48824

A range of novel expressed sequence tags (ESTs) associated with late developmental events during nodule organogenesis in the legume *Lotus japonicus* were identified using mRNA differential display; 110 differentially displayed polymerase chain reaction products were cloned and analyzed. Of 88 unique cDNAs obtained, 22 were unique to a nodule-specific expression pattern. The kinetics of mRNA accumulation of the majority of the ESTs analyzed were found to resemble the expression pattern observed for the *L. japonicus* leghemoglobin gene. These results indicate that the newly isolated molecular markers correspond to genes induced during late developmental stages of *L. japonicus* nodule organogenesis and provide important, novel tools for the study of nodule formation.

The formation of N-fixing nodules represents an unusual example of externally induced organogenesis that unites the plant host and symbiotic bacteria in a microenvironment appropriate for the support of bacterial N fixation and plant-mediated assimilation of N. This complex and highly regulated process begins with the specific recognition between the plant and bacterial partners, which leads to the synthesis of morphogenic lipochito-oligosaccharide molecules of bacterial origin, the Nod factors (for recent reviews, see Carlson et al., 1995; Spaink, 1996). Nodulation of legumes involves interactions between the plant and a specific group of *Rhizobium* species that can alter the expression of a number of host genes that function in N or C metabolism, proteins associated with the PBM, a family of oxyhemoproteins (leghemoglobins), and a number of proteins, the function of which remains to be elucidated (Delauney and Verma, 1988; de Bruijn and Schell, 1992; Mylona et al., 1995). A role for symbiotic bacteria (for a recent review, see Schulz et al., 1994). Specific plant genes called early nodulin genes are activated during early nodule morphogenetic events (Schulze et al., 1994; Mylona et al., 1995). Although the exact functions of the majority of the early nodulin genes identified to date have yet to be determined, the expression patterns of some of these genes have been correlated with processes such as preinfection (*Mtrp1*; Cook et al., 1995), infection (*ENOD3*, *ENOD12*, Scheres et al., 1990a, 1990b), and nodule meristem initiation or nodule structure formation (*cdc2-55*, *ENOD40*, *ENOD2*; Franssen et al., 1987; Miao et al., 1993; Crespi et al., 1994; Matvienko et al., 1994; van de Sande et al., 1996).

The successful invasion of plant root cells by symbiotic bacteria leads to the final stage of nodule formation, which culminates in the establishment of a fully developed N-fixing nodule. However, the developmental cues and molecular events underlying the late steps in nodule organogenesis are largely unknown. Several events that occur in both symbiotic partners at this late stage have been correlated with bacterial release from the infection thread and plant cell colonization (Sprent, 1989). These events include processes such as proliferation of the PBMs, bacteroid differentiation, and molecular and biochemical alterations that create and support the physiological environment required for N fixation and ammonia assimilation (Verma, 1992; Mylona et al., 1995).

Immediately prior to or concomitantly with the initiation of N fixation, a group of specific plant genes called late nodulin genes are activated (de Bruijn and Schell, 1992; Verma, 1992). Since these genes are not expressed in nodules lacking infected cells, it has been postulated that late nodulin genes may be coordinately expressed as a result of a single signal related to the release of bacteria from the infection thread (Nap and Bisseling, 1990; Welters et al., 1993). Typical members of this group include abundantly expressed genes encoding enzymes or subunits of enzymes that function in N or C metabolism, proteins associated with the PBM, a family of oxyhemoproteins (leghemoglobins), and a number of proteins, the function of which remains to be elucidated (Delauney and Verma, 1988; de Bruijn and Schell, 1992; Mylona et al., 1995). A role for

Abbreviations: DAL, days after inoculation; EST(s), expressed sequence tag(s); PBM, peribacteroid membrane.
small GTP-binding proteins (Rab1p and Rab7p homologs) in PBM development has also been proposed (Cheon et al., 1993), and a putative plant transcription factor, NMH7, present specifically in the infected bacteroid-containing cells, has been identified in alfalfa (Medicago sativa) root nodules (Heard and Dunn, 1995). It has been suggested that NMH7 may be involved in cellular activities specific to the differentiation of the infected cells (Heard and Dunn, 1995).

To study the complexity of genes expressed during the transition period between the development of the nodule structure and formation of a functional nodule, we examined the mRNA expression profiles in determinate nodules of the diploid legume Lotus japonicus (Handberg and Stougard, 1992; Cook et al., 1997; Jiang and Gresshoff, 1997) using the differential display method (Liang and Pardee, 1992). Here we describe the isolation of numerous ESTs corresponding to novel late nodulin genes, many of which appear to encode functions that have not previously been implicated in the nodule process.

MATERIALS AND METHODS

Plant Material

Lotus japonicus Gifu B-129-59 seeds (kindly provided by Dr. Jens Stougard, Aarhus University, Denmark) were surface-sterilized and germinated as described by Handberg et al. (1994). Rhizobium loti strain NZP2235 (Jarvis et al., 1982) was grown for 2 d at 28°C in TV medium (Atlas, 1997) and used to inoculate 7-d-old L. japonicus seedlings immediately before potting. For the initial experiments a 3:1 mixture of vermiculite:sand was used as plant growth medium. Later we refined the L. japonicus growing conditions, which resulted in faster and more efficient levels of nodulation (data not shown). Therefore, for the developmental northern-blot experiments shown in Figure 7, a 3:3:1 mixture of grade 2 vermiculite:sand was used. All plants were grown in a controlled environment in growth chambers (18-/6-h day/night cycle, 250 μE s⁻¹ m⁻² light intensity, 22/18°C day/night temperature, and 70% RH). B&D solution (Broughton and Dilworth, 1971) supplemented with 0.5 mM KNO₃ was used to water the plants. The relatively low concentration of combined N in B&D solution supports the growth of the uninoculated control plants but does not affect nodule formation on roots of infected L. japonicus plants (data not shown). Root segments were harvested at various times after infection, and fully developed nodules were collected 21 DAI. Plant tissues were immediately frozen in liquid N and stored at −70°C until use.

Nucleic Acid Isolation

R. loti genomic DNA was isolated following the procedure of Marmur (1961), except that the bacterial cells were washed with 1 M NaCl prior to pronase E digestion. For the isolation of plant genomic DNA, the method described by Rogers and Bendich (1988) was used.

Total plant RNA was isolated using the procedure of Verwoerd et al. (1989), except that the extraction buffer used was as described by Hall et al. (1978). The poly(A)⁺ fraction of mRNA was isolated using a minispin column kit (5 Prime-3 Prime, Boulder, CO) following the manufacturer’s instructions.

Developmental Differential Display of mRNA

The RNA differential display procedure was carried out using a mapping kit (RNAmap, GenHunter, Brookline, MA) following the manufacturer’s instructions (Goormachtig et al., 1995). The cDNA synthesis was performed using 0.5 μg of total RNA isolated from root segments (7, 11, and 13 DAI) and nodules (harvested 21 DAI). For control experiments, total RNA isolated from 7- and 21-d-old uninfected roots was used. Selected bands were reamplified and then blunt-ended by treatment with the Klenow fragment of DNA polymerase I (Boehringer Mannheim), phosphorylated at the 5‘ ends using T4 polynucleotide kinase, and cloned into the Smal-digested plasmid pK18 (Pridmore, 1987). Recombinant plasmids were mobilized into the Escherichia coli strain Invα-” (Novagen, Madison, WI) using electrot transformation. For each differentially displayed PCR fragment, 48 recombinant colonies were collected and stored individually in microtiter plates at −70°C in 20% glycerol.

Southern and Northern Analyses

For Southern analysis 10 μg of plant genomic DNA or 5 μg of R. loti total DNA was digested to completion using EcoRI endonuclease. The digested DNA was separated on a 0.8% agarose gel, transferred onto nitrocellulose filters (Hybond-N, Amersham), and UV cross-linked following standard procedures (Sambrook et al., 1989). Hybridizations and subsequent washes were carried out at 65°C using high-stringency conditions (Sambrook et al., 1989). For northern analysis 10 μg of total RNA or 4 μg of poly(A)⁺ RNA was separated on a 1.2% denaturing agarose gel (6% formaldehyde, 1× MOPS buffer: 20 mM Mops, 1.0 mM EDTA, 5.0 mM sodium acetate, pH 7.0) and transferred onto Nitro-Plus membranes (Fisher Scientific) as described previously (Sambrook et al., 1989). Ten micrograms of total RNA was used for the slot-blot RNA hybridization. Prehybridizations and hybridizations were performed according to the procedure described by Church and Gilbert (1984). Filters were washed twice for 15 min in 2× SSC and 0.1% SDS, once for 15 min in 0.5× SSC and 0.1% SDS, and once for 10 min in 0.1× SSC, at 65°C. DNA probes were labeled with [α-32P]dATP using the random-priming kit from Boehringer Mannheim following the manufacturer’s instructions.

Differential Colony Hybridization

For differential colony hybridization microtiter “combo” plates were developed. A single 96-well microtiter plate combined representatives of eight (A–H) differentially displayed PCR products, each represented by 12 recombinant
RESULTS

Molecular Characterization of *L. japonicus* Root Nodule Development

Upon infection with *R. loti*, *L. japonicus* plants form determinate nodules (Handberg and Stougaard, 1992; Jiang and Gresshoff, 1993); however, a considerable amount of variation in nodulation efficiency has been observed with different *R. loti* strains (data not shown). Therefore, to develop an efficient system we examined the nodulation characteristics of 21 different *R. loti* strains (provided by Dr. D.B. Scott, Massey University, New Zealand). Out of this collection, a highly efficient *R. loti* strain, NZP2235, was selected and used for further nodulation of *L. japonicus* ecotype Gifu plants (K. Szczyglowski and F.J. de Bruijn, unpublished data). Under the plant growth conditions used in this study, visible signs of nodule formation could be detected on the roots of *L. japonicus* plants between 3 and 5 DAI. During the next few days, usually between 7 and 11 DAI, nodule structures enlarged and turned pink, indicating leghemoglobin synthesis.

To further define the early and late stages of *L. japonicus* nodule development, RNA gel-blot analyses were performed with two well-characterized marker genes, the *L. japonicus* *Enod2* gene (R. Chen, K. Szczyglowski, and F.J. de Bruijn, unpublished data) and a cDNA corresponding to an *L. japonicus* leghemoglobin gene (J. Stoltzfus, K. Szczyglowski, and F.J. de Bruijn, unpublished data). In a number of legume species, the expression of the early nodulin gene *Enod2* and the leghemoglobin genes has been correlated with early and late stages of nodule development, respectively (Nap and Bisseling, 1990; Mylona et al., 1995).

Developmental northern analysis using total RNA isolated from roots and nodules at different times after *L. japonicus* inoculation revealed that the *Enod2* gene was expressed at a very low level in the uninfected control roots (Fig. 1). The level of *Enod2* mRNA gradually increased between 7 and 21 DAI. In comparison, the expression of the *L. japonicus* leghemoglobin gene was first detected 11 DAI. Based on these observations we concluded that the late developmental events in the *L. japonicus* nodule morphogenesis process are likely to occur between 7 and 11 DAI. Therefore, this period became the focus of our further molecular analysis.

![Figure 1. RNA gel-blot analysis of *L. japonicus* *LjEnod2* and leghemoglobin (*Ljglb1*) gene expression. Ten micrograms of total RNA isolated from 7-d-old uninfected roots (control) and root segments or nodules harvested 7, 11, 13, and 21 DAI were analyzed. α-32P-labeled, single-stranded cDNAs encoding *L. japonicus* *LjEnod2* and *Ljglb1* were used as molecular probes.](www.plantphysiol.org)
Construction of a *L. japonicus* Nodule-Specific EST Library

The strategy used to construct the *L. japonicus* nodule-specific EST library was based on mRNA differential display fingerprinting (Liang and Pardee, 1992). We applied this procedure to detect and clone transcripts that were specifically expressed during the transition period between the formation of the nodule structure and the onset of N fixation. To minimize false-positive results (Liang et al., 1993) and to maximize the probability of isolating nodule-specific ESTs, we compared the RNA profiles derived from four relatively late phases of nodule development. Using the four degenerate T12MN primers in combination with 20 arbitrary decamers, we displayed RNA species from the *L. japonicus* roots harvested 7, 11, and 13 DAI and from 21-d-old nodules and compared them with the RNA profiles derived from 7- and 21-d-old uninfected control roots. Each primer set was found to generate 80 to 150 bands. Since a total of 80 primer combinations were used, approximately 10,000 PCR products were displayed for every time point analyzed. A representative example of these experiments is shown in Figure 2.

A visual inspection of all RNA profiles obtained revealed that approximately 1.4% of the bands (137 of 10,000) appeared to be present in infected *L. japonicus* roots and/or nodules but not in control, uninfected roots (Fig. 2). Of 137 bands recovered from the polyacrylamide gels, 110 bands were successfully reamplified and cloned into the pK18 vector (Pridmore, 1987). Given the likelihood that the reamplified PCR products represented a mixture of different mRNA species (Bauer et al., 1993), 48 recombinant colonies were stored per individual PCR product in a single microtiter plate. In addition, 8 PCR products, each represented by 12 recombinant colonies, were combined in the 96-well microtiter combo plate and used for differential hybridization.

Differential Hybridization and Identification of Nodule-Specific ESTs

Replica filters from each of the 14 microtiter combo plates, representing a total of 110 PCR bands, were differentially hybridized with radiolabeled cDNAs derived from nodule or root poly(A)⁺ RNA. This screening procedure permitted the selection of a group of recombinant colonies, representing 39 differentially displayed PCR products that hybridized specifically with the nodule-derived cDNA probe but not with the probe from uninfected root. Twenty-six colonies representing 7 different PCR bands hybridized with probes from both uninfected roots and nodules, whereas 11 colonies corresponding to 6 PCR products hybridized only to root cDNAs.

The recombinant colonies associated with the remaining 58 PCR bands did not show detectable signals with either probe. An example of this differential hybridization analysis is shown in Figure 3. In this example, several different outcomes of the differential hybridization procedure are shown. The top two rows of nodule-specific strong hybridization signals in Figure 3B (top filter), representing a group of LjN50 cDNAs, illustrate the specificity of the differential hybridiza-

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Developmental mRNA differential display. mRNA amplification profiles from 21-d-old (lanes 1) and 7-d-old (lanes 2) uninfected, control roots and root segments and nodules harvested 7 (lane 3), 11 (lanes 4), 13 (lanes 5), and 21 (lanes 6) DAI. The decamer primer AP9 (CGTGCGCAATA), in combination with four 3' anchoring primers (T₁₂MG, T₁₂MA, T₁₂MT, and T₁₂MC), was used to generate the RNA profiles shown. The dots indicate putative nodule-specific bands. The differentially displayed band generated using the AP9/T₁₂MC primer combination (lanes 6) corresponds to the LjN77 EST and encodes an *L. japonicus* leghemoglobin (Table I). Nt, Nucleotides.

all) individually picked transformants containing the same cDNA that most likely corresponds to highly expressed nodulin genes. The next two rows and the last two rows on the top filter show cases in which no signal was found with either nodule- or root-specific RNA, suggesting that the cDNAs contained in these colonies either represent false positives or correspond to rare mRNA species. A more sensitive approach was used to establish their tissue specificity (see “Discussion”). The third set of rows on the top filter show isolated cDNAs that do not appear to be nodule specific. The examples shown on the bottom filter show a limited number of positive cDNAs among the transformants analyzed (LjN77, LjN50). These results directly show the need to characterize several independent colonies for each putative cDNA. Based on the differential hybridiza-
protein sequences stored in the databases. The translated LjN clones revealed significant homology to other DNA/protein alignments for LJN13 are shown in LjNIOl dase gene (Madsen et al., 1993; Yang et al., 1993). The DNA coproporphyrinogen oxidase gene of Glycine max (Ahmad et al., 1993) and the amino acid sequence of LjN3 showed strong similarity to the Arabidopsis thaliana protein phosphatase 2C (Kumon and Yama-

Figure 4, B and C, respectively. In addition, seven different leghemoglobin cDNA sequence isolated from a nodule-specific cDNA library and used for the initial northern analysis (data not shown). The majority of the LjN cDNAs were found to contain DNA sequences of the same mRNA species was amplified from a slot-blot northern analyses (Table I). The full-copy cDNAs corresponding to LjN53 and LjN63 were isolated and predicted to encode a 65-kD protein (Nlj65) and a glutamate-rich (27%) protein of 192 amino acids (Nlj21). The deduced amino acid sequence of Nlj65 showed significant similarity with the peptide transporter AtPTR2-B from Arabidopsis and other species (Frommer et al., 1994; Fig. 5B). The deduced amino acid sequence of nodulin Nlj21 shared substantial similarity with the Agl3 protein from Alnus glutinosa nodules (Guan et al., 1997) and the pKiWIS01 gene specifically expressed during fruit development of kiwifruit (Actinidia deliciosa) (Ledger and Gardner, 1994; Fig. 6). The expression of the ag13 gene has been localized to the pericycle of the vascular bundle of A. glutinosa nodules and in infected cells that exhibit degradation of the endosymbiont (Guan et al., 1997).

The LjN93, LjN5, and LjN48 cDNAs were found to share significant similarity to as-yet-uncharacterized cDNA/EST sequences from different plant species (Table I). The full-copy cDNA corresponding to the L. japonicus LjN5 cDNA was isolated and shown to encode a 15.6-kD protein (data not shown). The expression of the corresponding gene was induced in infected cells of L. japonicus at approximately the time of initiation of N fixation, and the deduced protein was found to share significant similarity with predicted α-helical domains of two related anonymous Arabidopsis ESTs (Kapranov et al., 1997).

The remaining seven nodule-specific LjN cDNAs did not show any significant similarity to DNA/protein sequences stored in the databases. Thus, of 19 unique LjN clones isolated, 14 are likely to represent novel nodulin genes.

Expression Analysis of the LjN ESTs

To study the temporal expression pattern of selected L. japonicus ESTs, developmental slot-blot northern analyses were performed. Total RNA isolated from uninfected L. japonicus control roots and root segments or from nodules infected with A. castellanii (control) or A. aphrophos (roots were hybridized with a
representative selection of radiolabeled cDNA inserts (Fig. 7). Since the plant material used in this experiment was generated using slightly modified growth conditions (see "Materials and Methods"), the LjN77-derived insert, encoding one of the L. japonicus leghemoglobins, was used as a marker gene for late developmental events (Table I; Fig. 7). The mRNA species corresponding to the L. japonicus leghemoglobin genes were first detectable 7 DAI, which was slightly earlier than the leghemoglobin gene expression pattern obtained during our initial experiments (compare Figs. 1 and 7). The corresponding mRNA accumulated gradually to a high level in 21-d-old nodules.

All LjN cDNA inserts analyzed hybridized in a nodule-specific or enhanced manner, except LjN3, which did not give a clearly detectable signal with either root- or nodule-derived total RNA (see below). The expression of some of the genes analyzed could also be detected at a very low level in the uninfected roots (e.g. LjN 101, LjN71, LjN93, and LjN73). However, a significant increase in the level of their corresponding mRNAs was apparent at approximately 7 DAI, which was similar to the expression patterns of the other L. japonicus ESTs analyzed (Fig. 7).

The expression pattern of the majority of genes analyzed resembled that of the leghemoglobin gene (LjN77), indicating that they are likely to represent L. japonicus late nodulin genes. mRNA of the coproporphyrinogen oxidase homolog LjN101 accumulated in a slightly different fashion than the mRNA of the leghemoglobin gene LjN77. The gene corresponding to LjN101 appeared to be induced to a high level at approximately 7 DAI, with no significant changes in the steady-state level of mRNA accumulation during the later stages of nodule development (Fig. 7).

The mRNA corresponding to the LjN13, an ENOD40 homolog (Table I; Fig. 4B), was weakly detectable in uninfected roots, but the hybridization signals were significantly enhanced in infected roots and fully developed 21-d-old nodules. The poly(A) fraction of total RNA from uninfected roots and nodules was used to further analyze tissue-specific expression of LjN3, the putative protein phosphatase 2C homolog. The LjN3 insert hybridized with both root and nodule mRNA species of approximately 1600 nucleotides in length (Fig. 8). However, the level of the corresponding mRNA in L. japonicus nodules was found to be approximately 6 times higher than in uninfected roots, confirming the nodule-specific/enhanced pattern of LjN3 gene expression.

Since all EST sequences analyzed were generated using PCR-based procedures, their plant (L. japonicus) origin needed to be confirmed by Southern analysis. All LjN cDNA inserts hybridized specifically with L. japonicus genomic DNA but not with total DNA isolated from R. loti strain NZP2235 (data not shown).

### Table I. Sequence similarities detected for the nodule-specific LjN clones

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Accession No.</th>
<th>No. of Isolates</th>
<th>Size Range</th>
<th>Best Homology</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>LjN3</td>
<td>AF000382</td>
<td>1</td>
<td>359 bp</td>
<td><em>A. thaliana</em> protein phosphatase 2C (Kuromoni and Yamamoto, 1994)</td>
<td>1.4 x 10^-19</td>
</tr>
<tr>
<td>LjN13</td>
<td>AF000383</td>
<td>3</td>
<td>372 bp</td>
<td><em>G. max</em> mRNA ENOD40-2 (Yang et al., 1993)</td>
<td>6.9 x 10^-27d</td>
</tr>
<tr>
<td>LjN53</td>
<td>AF000392b</td>
<td>1</td>
<td>167-170 bp</td>
<td><em>M. sativa</em> leghemoglobin (Lobler and Hirsch, 1992)</td>
<td>3.0 x 10^-6d</td>
</tr>
<tr>
<td>LjN63</td>
<td>AF000402b</td>
<td>1</td>
<td>415 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>7.6 x 10^-27d</td>
</tr>
<tr>
<td>LjN73</td>
<td>AF000404</td>
<td>2</td>
<td>297-300 bp</td>
<td><em>Canavalia erecta</em> leghemoglobin mRNA (accession no. U09671)</td>
<td>3.0 x 10^-11d</td>
</tr>
<tr>
<td>LjN77</td>
<td>AF000405</td>
<td>5</td>
<td>162-570 bp</td>
<td><em>Canavalia erecta</em> leghemoglobin mRNA (accession no. U09671)</td>
<td>2.1 x 10^-16</td>
</tr>
<tr>
<td>LjN132</td>
<td>AF000390</td>
<td>2</td>
<td>162-570 bp</td>
<td><em>Canavalia erecta</em> leghemoglobin mRNA (accession no. U09671)</td>
<td>1.8 x 10^-14</td>
</tr>
<tr>
<td>LjN36</td>
<td>AF000406</td>
<td>1</td>
<td>228-238 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>4.4 x 10^-19</td>
</tr>
<tr>
<td>LjN101</td>
<td>AF000407</td>
<td>1</td>
<td>228-238 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>7.6 x 10^-27d</td>
</tr>
<tr>
<td>LjN5</td>
<td>AF000404</td>
<td>9</td>
<td>483 bp</td>
<td><em>Canavalia erecta</em> leghemoglobin mRNA (accession no. U09671)</td>
<td>NS</td>
</tr>
<tr>
<td>LjN71</td>
<td>AF000387</td>
<td>1</td>
<td>607 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>NS</td>
</tr>
<tr>
<td>LjN80</td>
<td>AF000386</td>
<td>2</td>
<td>415 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>NS</td>
</tr>
<tr>
<td>LjN81</td>
<td>AF000385</td>
<td>2</td>
<td>322-462 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>NS</td>
</tr>
<tr>
<td>LjN112</td>
<td>AF000384</td>
<td>1</td>
<td>228-238 bp</td>
<td><em>Ricinus communis</em> cDNA (van de Loo et al., 1995)</td>
<td>NS</td>
</tr>
</tbody>
</table>

* a Probability that such a match would occur merely by chance as given by BLAST. b Full-copy or almost full-copy cDNA corresponding to the indicated EST has been deposited to GenBank under this accession number. c NS, No significant match. d DNA homology.
Further Characterization of the *L. japonicus* EST Library

To further characterize the *L. japonicus* EST library, we used automated DNA sequencing. Two randomly selected recombinant colonies, both corresponding to a single differentially displayed PCR product, were used for DNA-sequence analysis. The recombinant colonies were selected from the collection of *L. japonicus* EST combo plates based on their lack of hybridization with the radiolabeled root and nodule cDNA probes used during the differential hybridization procedure. The nucleotide sequence of 142 cDNA inserts representing a total of 71 PCR products was established. DNA sequence comparisons showed that they corresponded to 69 unique cDNA sequences (data not shown).

The EST nucleotide sequences obtained were compared with the nucleotide and protein sequences in the databases by BLASTN and BLASTX searches, respectively (Altschul et al., 1990). Nine of 69 EST sequences analyzed showed a moderate to high level of similarity to DNA/protein sequences stored in the databases. A summary of the results of this analysis is shown in Table II. We refer to these clones as *L. japonicus* *L* EST cDNAs. All cDNAs except *L*ESTS9 appeared to encode unique enzymatic functions, including subtilisin-like protease (Ribeiro et al., 1995), adenosylsucinate synthetase (Zeidler et al., 1992), tyr dehydration (Yenofsky et al., 1988), dehydroquinate dehydratase carboxylase (Maldonado-Mendoza et al., 1996), lipoxygenase (Shibahara et al., 1985), chalcone synthase to sensitive protocols, such as reverse transcription-PCR. The remaining 60 EST sequences obtained via the random-sequencing approach did not show any significant homologies or similarities and were therefore classified as anonymous *L. japonicus* ESTs.

**DISCUSSION**

We have previously reported the successful application of the mRNA differential display technique for the identification and isolation of early nodulin genes from *Sesbania rostrata* stem nodules (Goormachtig et al., 1995). In the present study we used the reverse transcription-PCR-based differential display procedure to identify a range of molecular markers associated with relatively late developmental events during *L. japonicus* nodule organogenesis. A well-characterized nodulin gene, the leghemoglobin gene, was used to define the late stages in *L. japonicus* nodule development. Using 80 primer combinations, we analyzed the profiles of approximately 10,000 PCR products and we successfully reamplified and cloned 110 differentially displayed nodule-specific or enhanced bands. Thus, a library of differentially displayed *L. japonicus* ESTs was established. The differential hybridization procedure revealed relatively abundant mRNA species and allowed the selection of 39 nodule-specific PCR products representing 19 unique cDNA sequences. For the purpose of this paper, the term “specific” was used merely to reflect the significant differences in the expression levels observed between the uninfected roots and infected roots or nodules.
Figure 6. Amino acid sequence comparison of Nlj21, Ag13, and PkWiW1501. The deduced amino acid sequence of the L. japonicus nodulin Nlj21 is compared with A. glutinosa Ag13 and kiwifruit PkWiW1501 (Guan et al., 1997; Ledger and Gardner, 1994). The identical residues and conservative substitutions are boxed.

The apparent redundancy found among some of the isolated sequences (between 2 and 9, see Table I) was in part due to the different combinations of primers used during the differential display procedure. In most cases analyzed different positions of \( T_3 \)MN primers at the 3' end of the mRNA, in combination with a single arbitrary decamer primer, gave rise to multiple products derived from the same mRNA species (e.g. LjN3, LjN101, and LjN81, Table I). This finding may be explained by the presence of multiple poly(A) sites in a given plant mRNA species (Wu et al., 1995). In the case of the LjN50 cDNA group, in addition to differences in the position of the \( T_3 \)MN primers, three different arbitrary decamers contributed to multiple independent isolations of the respective cDNA species (data not shown). The sequence analysis of nine partial cDNAs belonging to the LjN50 group revealed that all of them shared highly conserved 3' terminal sequences of approximately 220 bp. However, significant differences in the corresponding DNA sequences at the 5' ends were found, indicating that they may correspond to related, but not identical, genes in the L. japonicus genome (data not shown).

The comparative sequence analysis of representatives of all 19 cDNA groups established that the majority of them are likely to correspond to novel nodule-specific genes. This group of L. japonicus ESTs includes putative homologs of protein phosphatase 2C and Cyt P450, both candidates for proteins with regulatory functions. The protein phosphatase activity in plants, of which the least well-understood is the presence of multiple poIy(A) sites in a given plant mRNA from the same mRNA species (e.g. LjN3, LjN101, and LjN81, Table I). This finding may be explained by the presence of multiple poly(A) sites in a given plant mRNA species (Wu et al., 1995). In the case of the LjN50 cDNA group, in addition to differences in the position of the \( T_3 \)MN primers, three different arbitrary decamers contributed to multiple independent isolations of the respective cDNA species (data not shown). The sequence analysis of nine partial cDNAs belonging to the LjN50 group revealed that all of them shared highly conserved 3' terminal sequences of approximately 220 bp. However, significant differences in the corresponding DNA sequences at the 5' ends were found, indicating that they may correspond to related, but not identical, genes in the L. japonicus genome (data not shown).

The comparative sequence analysis of representatives of all 19 cDNA groups established that the majority of them are likely to correspond to novel nodule-specific genes. This group of L. japonicus ESTs includes putative homologs of protein phosphatase 2C and Cyt P450, both candidates for proteins with regulatory functions. The protein phosphatase activity in plants, of which the least well-understood is the presence of multiple poIy(A) sites in a given plant mRNA species (e.g. LjN3, LjN101, and LjN81, Table I). This finding may be explained by the presence of multiple poly(A) sites in a given plant mRNA species (Wu et al., 1995). In the case of the LjN50 cDNA group, in addition to differences in the position of the \( T_3 \)MN primers, three different arbitrary decamers contributed to multiple independent isolations of the respective cDNA species (data not shown). The sequence analysis of nine partial cDNAs belonging to the LjN50 group revealed that all of them shared highly conserved 3' terminal sequences of approximately 220 bp. However, significant differences in the corresponding DNA sequences at the 5' ends were found, indicating that they may correspond to related, but not identical, genes in the L. japonicus genome (data not shown).

The comparative sequence analysis of representatives of all 19 cDNA groups established that the majority of them are likely to correspond to novel nodule-specific genes. This group of L. japonicus ESTs includes putative homologs of protein phosphatase 2C and Cyt P450, both candidates for proteins with regulatory functions. The protein phosphatase activity in plants, of which the least well-understood is the presence of multiple poIy(A) sites in a given plant mRNA species (e.g. LjN3, LjN101, and LjN81, Table I). This finding may be explained by the presence of multiple poly(A) sites in a given plant mRNA species (Wu et al., 1995). In the case of the LjN50 cDNA group, in addition to differences in the position of the \( T_3 \)MN primers, three different arbitrary decamers contributed to multiple independent isolations of the respective cDNA species (data not shown). The sequence analysis of nine partial cDNAs belonging to the LjN50 group revealed that all of them shared highly conserved 3' terminal sequences of approximately 220 bp. However, significant differences in the corresponding DNA sequences at the 5' ends were found, indicating that they may correspond to related, but not identical, genes in the L. japonicus genome (data not shown).
cDNA clones that share significant similarity to already characterized late nodulin genes from different legume species were also identified. The latter group includes *L. japonicus* leghemoglobin genes and a putative homolog of soybean coproporphyrinogen oxidase, an enzyme catalyzing the oxidative decarboxylation of coproporphyrinogen III to protoporphyrinogen IX in the heme and chlorophyll biosynthetic pathway(s) (Madsen et al., 1993). The latter finding supports the notion that specific plant host functions participate in the synthesis of heme, which is needed in nodules for increased hemoprotein biosynthesis (Madsen et al., 1993).

The kinetics of mRNA accumulation of the *L. japonicus* ESTs analyzed closely resemble the pattern observed for leghemoglobin gene expression in *L. japonicus* nodules, indicating that a majority of them are almost certainly related to late stages in nodule development. Since the goal of our analysis was to identify molecular markers associated specifically with late developmental stages in nodule organogenesis, our results illustrate the advantage of using nodules with a determinate developmental pattern for late nodulin expression analysis as opposed to indeterminate nodules, in which successive early and late developmental stages coexist.

Since a majority of the *L. japonicus* EST clones did not hybridize with the root- or nodule-specific cDNA probes used during differential hybridization screening, we used a random-sequencing approach to further characterize the remaining differentially displayed *L. japonicus* ESTs. This resulted in the generation of 69 unique cDNA sequences. Only 9 of the cDNAs shared similarity with protein sequences from the databases (Table II). Since selection of all of these sequences was based exclusively on the observed differential display patterns, it is not certain whether they indeed correspond to nodule-specific or nodule-enhanced genes. Clearly, more in-depth analyses are needed to unambiguously establish their tissue or cell specificity. However, it is important to note that at least three of them, a subtilisin-like protease, chalcone reductase, and Gin phosphoribosylpyrophosphate amidotransferase, have previously been reported to be induced during nodulation (Goormachtig et al., 1995; Kirn et al., 1995; Ribeiro et al., 1995). Several roles for lipoxygenases in plant-microbe interactions, including symbiotic nodule formation, have also

transduction, hormonal regulation, mitosis, and control of C and N metabolism (Smith and Walker, 1996). Cyt P450 enzymes, on the other hand, are membrane-bound, heme-containing enzymes that are implicated in a variety of biosynthesis reactions (Nelson et al., 1993; Frey et al., 1995). In plants they are typically involved in defense mechanisms and in the synthesis of chemically diverse secondary metabolites, plant hormones, or cell wall-related substances (Holton et al., 1993; Frey et al., 1995; Winkler and Helentjaris, 1995; Szekeres et al., 1996). The specific functions of these interesting new nodulin genes (*LjN3* and *LjN73*) in *L. japonicus* nodules remain to be elucidated.
been suggested (Croft et al., 1993; Gardner et al., 1996; Perlick et al., 1996; Veronesi et al., 1996).

In summary, of 110 L. japonicus differential amplification products, 88 unique partial cDNA sequences were obtained. Nineteen of these were further analyzed and were shown to correspond to nodule-specific genes, with the majority of them likely to encode novel nodule-specific functions. Twenty-two L. japonicus ESTs showed varying degrees of similarity to different DNA/protein sequences stored in the databases. For most of these, solid predictions about their activities could be made. However, their specific roles in the late stages of nodule development. The collection of ESTs reported here will also be an indispensable tool for the development of an L. japonicus genetic map, an essential step toward future map-based cloning of symbiosis-specific genes in this model legume plant.

ACKNOWLEDGMENTS

We thank Brian MacSpadden Gardner for comments and helpful discussions. We also thank Kurt Stepnitz for help with preparing the figures.

Table II. Sequence similarities detected for randomly sequenced EST clones

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>Accession No.</th>
<th>No. of Isolates</th>
<th>Size</th>
<th>Best Homology</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>LjEST38</td>
<td>AF000393</td>
<td>1</td>
<td>411</td>
<td>A. thaliana subtilisin-like protease (Ribeiro et al., 1995)</td>
<td>DNA/protein 3.8 x 10^-10</td>
</tr>
<tr>
<td>LjEST58</td>
<td>AF000394</td>
<td>1</td>
<td>527</td>
<td>Yeast adenosylsucinate synthetase (Zeidler et al., 1993)</td>
<td>2.7 x 10^-7</td>
</tr>
<tr>
<td>LjEST59</td>
<td>AF000395</td>
<td>1</td>
<td>258</td>
<td>Unknown Trypanosoma brucei protein (accession no. U05313)</td>
<td>5.4 x 10^-8</td>
</tr>
<tr>
<td>LjEST66</td>
<td>AF000396</td>
<td>1</td>
<td>245</td>
<td>Papaver somniferum tyrosine decarboxylase (Maldonado-Mendoza et al., 1996)</td>
<td></td>
</tr>
<tr>
<td>LjEST95</td>
<td>AF000397</td>
<td>1</td>
<td>254</td>
<td>G. max lqo3 gene (Yenofsky et al., 1988)</td>
<td>8.3 x 10^{-9a}</td>
</tr>
<tr>
<td>LjEST103</td>
<td>AF000398</td>
<td>1</td>
<td>131</td>
<td>Nicotiana tabacum dehydrogenase/shikimate dehydrogenase mRNA 3' end (Booner and Jensen, 1994)</td>
<td>1.8 x 10^{-12a}</td>
</tr>
<tr>
<td>LjEST105</td>
<td>AF000399</td>
<td>1</td>
<td>238</td>
<td>Rat heme oxygenase (Shibahara et al., 1985)</td>
<td>3.4 x 10^{-5}</td>
</tr>
<tr>
<td>LjEST118</td>
<td>AF000400</td>
<td>1</td>
<td>296</td>
<td>S. rostrata chalcone reductase (Goormachtig et al., 1995)</td>
<td>1.7 x 10^{-11}</td>
</tr>
<tr>
<td>LjEST120</td>
<td>AF000401</td>
<td>1</td>
<td>192</td>
<td>G. max glutamine phosphoribosyl-pyrophosphate amidotransferase (Kim et al., 1995)</td>
<td>8.0 x 10^{-8}</td>
</tr>
</tbody>
</table>

* DNA homology.

LITERATURE CITED


Cheon C-III, Lee N-G, Siddique A-BM, Bal AK, Verma DPS (1993) Roles of plant homologs of Rab1p and Rab7p in the biogenesis of the peribacteroid membrane, a subcellular compartment formed de novo during root nodule symbiosis. EMBO J 12: 4125-4135


Croft KPC, Juttner F, Sliusarenko AJ (1993) Volatile products of the lipoygenase pathway evolved from Phaseolus vulgaris (L.)
leaves inoculated with *Pseudomonas syringae pv phaseoli.*

Plant Physiol 101: 13–24


