Thirteen nodule-specific or nodule-enhanced genes encoding products homologous to cysteine cluster proteins or plant lipid transfer proteins are identified in *Astragalus sinicus* L. by suppressive subtractive hybridization

Min-Xia Chou, Xin-Yuan Wei, Da-Song Chen and Jun-Chu Zhou*

State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, 430070 Wuhan, People’s Republic of China

Received 13 December 2005; Accepted 17 April 2006

Abstract

Thirteen nodule-specific or nodule-enhanced genes have been revealed by suppressive subtractive hybridization (SSH) with two mRNA populations of infected and uninfected control roots of *Astragalus sinicus*. Eleven of them encode small polypeptides showing homology to cysteine cluster proteins (CCPs) that contain a putative signal peptide and conserved cysteine residues. Among these CCP-like genes, *AsG257* codes for a homologue of the defensin 2 family and *AsD255* contains a scorpion toxin-like domain at the C-terminus. Sequence analysis of a genomic *AsD255* fragment which was isolated revealed that one intron separates the first exon encoding the signal peptide from the second exon encoding the cysteine cluster domain of this nodulin. Another two genes, *AsE246* and *AsIB259*, encode two different products similar to lipid transfer proteins (LTPs). Virtual northern blot and reverse transcription–polymerase chain reaction (RT–PCR) analysis indicated that the other genes except *AsIB259* and *AsC2411* were expressed exclusively in inoculated roots and that their expression was 2–4 d later than that of the leghaemoglobin (Lb) gene during nodule development. Transcription of *AsIB259* was also detected in uninfected control roots but with a significant decline in expression and a temporal expression similar to *Lb*. *AsC2411* had a basal expression in control roots identified by RT–PCR. Sequence alignment showed that the putative proteins *AsE246* and *AsIB259* show lower homology with LTPs from legumes than with those from other plants.

Key words: *Astragalus sinicus* L, cysteine cluster protein, lipid transfer protein, nodulin, suppressive subtractive hybridization.

Introduction

Legumes have established symbiotic associations with soil micro-organisms such as rhizobia and fungi to obtain mineral nutrition. The symbiosis between leguminous plants and rhizobia leads to the formation of a novel root organ, the nodule. In mature nodules, nitrogen-fixing rhizobia can supply the host plant with ammonium. Two major classes of nodules are developed on legume roots: the indeterminate type and the determinate type. Indeterminate nodules, usually formed by temperate legumes, display a characteristic zonation. The apical meristem is zone I, followed by infection zone II where post-meristematic cells gradually differentiate and become infected with rhizobia. The major transcriptional changes occur in the amyloplast-rich interzone II–III. The nitrogen-fixing zone III is filled with bacteroids enclosed by peribacteroid membrane derived from the plant, and the last zone is the senescent zone IV (Vasse *et al*., 1990; Brewin, 1991; Hirsch, 1992). Determinate nodules are developed on tropical legumes such as *Mimosa* and *Dioclea*, which are usually formed by tropical legumes.
nODULES are commonly formed on the roots of tropical legumes such as *Lotus japonicus*, *Glycine*, or *Vigna*. In this class of nodules, hypodermic cell division only persists for several days and subsequent growth is completed by cell expansion. The result of this is that the central tissue containing nitrogen-fixing cells is consistent and there is no zonation in mature nodules (Brewin, 1991; Hirsch, 1992).

In recent years, dozens of nodulin genes have been isolated from various leguminous plants, especially the model legumes *Medicago truncatula* and *Lotus japonicus*. To date, nearly 700 000 nucleotide sequences representing the Fabaceae are available from the National Center for Biotechnology Information (Graham et al., 2004), and functional genomic studies have been carried out on the model legumes *M. truncatula* and *L. japonicus*, and on the crop legume soybean (VandenBosch and Stacey, 2003). By screening sequences homologous to cytosine cluster proteins (CCPs) identified by Györgyey et al. (2000) against the *M. truncatula* expressed sequence tags (ESTs), Mergaert et al. (2003) detected 311 distinct tentative consensus sequences (TCSs) or singletons belonging to a gene family named NCR (nodule-specific cytoine-rich). The NCR polypeptides are very small (60–90 amino acids) and highly divergent, with the exception of a conserved signal peptide and a conserved cytoine motif with a characteristic arrangement of four, six, eight, or 10 cysteine residues probably involved in disulphide bridges at the C-terminal domain. Gene expression profiles have revealed that they are all nodule specific and exhibit distinct temporal and spatial expression patterns coupled to different stages of development from nodule primordial to senescent nodules. It has been shown by green fluorescent protein fusions expressed in onion (*Allium cepa*) epidermal cells that the signal peptide targets the polypeptides in the secretory pathway. A concerted action in nodule development has been suggested by the co-regulation of certain pathway. A concerted action in nodule development has been suggested by the co-regulation of certain pathway.

The result of this is that the central tissue containing nitrogen-fixing cells is consistent and there is no zonation in mature nodules (Brewin, 1991; Hirsch, 1992).

Although so many nodulin genes have been reported, it is still unclear about the exact roles of most of them, just as for the CCP genes described above. More importantly, it remains undefined how much information about other legumes, especially legume crops, can be acquired directly from legume models (VandenBosch and Stacey, 2003).

*Astragalus sinicus* L. (Chinese milk vetch) is one of the most popular legumes, widely grown as a green-manure-forage in Asian countries including China, Japan, and Korea. In addition, it is used as one of the important Chinese traditional medicines. *Astragalus sinicus* has a specific symbiotic relationship with the rhizobia of *Mesorhizobium huakuii* which forms indeterminate-type nodules (Chen et al., 1991). Also, it sets up symbiosis with the endomycorrhizal fungus *Glomus intraradices*. *Astragalus sinicus* is a very specific host plant and usually forms nodules only with rhizobia isolated from itself (Chen and Shu, 1944). The only reported exception of cross-inoculation was with a rhizobial strain isolated from *Astragalus ciceri* (Malek et al., 1998). To work with a smaller size than *M. truncatula* and shorter generation time, the interaction between Chinese milk vetch and *M. huakuii* can be studied by using a test tube nodulation system. To date, there is very little information on the involvement of nodulation in the Chinese milk vetch. More than 100 nodule-specific or nodule-enhanced cDNA clones were isolated from Chinese milk vetch by differential display, and a novel nodulin gene, *AsNODe22*, coding for an 18 kDa protein with unknown function, was identified (Fujie et al., 1998). Subsequently, Naito et al. (2000) reported another nodule-specific gene *AsNODf32* encoding a polypeptide similar to cytoine proteinas.

To identify the genes involved in nodule formation and to increase usable molecular probes, a cDNA library of *A. sinicus* genes specifically expressed in infected roots was generated using a polymerase chain reaction (PCR)-based suppressive subtractive hybridization (SSH) technique in this study. Thirteen nodule-specific or nodule-enhanced genes have been identified and their expression patterns are presented. Among the 11 genes encoding CCP-like proteins, the putative *AsG257* polypeptide carries a defensin 2-like protein and a conserved cysteine motif with a characteristic arrangement of four, six, eight, or 10 cysteine residues for a signal peptide peptidase. Besides galegoid legumes forming indeterminate nodules such as *M. truncatula*, pea (Kato et al., 2002), broad bean (Frühling et al., 2000), white clover (Crockard et al., 2002), and *Galega orientalis* (Kajialainen et al., 2002), CCP homologues were also identified in the soybean genome with hybridization of *M. truncatula* nodule-specific CCPs to soybean bacterial artificial chromosomes (BACs) by Graham et al. (2004). However, the same authors failed to identify nodule-specific CCPs from soybean or *L. japonicus* forming determinate nodules.

Materials and methods

Plant growth and inoculation

Seeds of *A. sinicus* were surface-sterilized by 5 min treatment with 95% (v/v) ethanol, 10 min with 5% (v/v) NaClO, and eight washes with axenic water, and were then germinated at room temperature in the dark for 2 d.

The germinated seeds were transferred to sterilized sand pots, fertilized with Fahraeus nitrogen-free nutrient solution, and cultivated in a greenhouse with a 16 h light/8 h dark cycle at 18–22 °C. Six-d-old seedlings were inoculated with *M. huakuii* 7653R.

RNA and DNA preparation

For SSH, virtual northern blotting, and rapid amplification of cDNA ends (RACE), RNA was isolated from inoculated and non-inoculated 10-old seedlings were inoculated with Fahraeus nitrogen-free nutrient solution, and cultivated in a greenhouse with a 16 h light/8 h dark cycle at 18–22 °C. Six-d-old seedlings were inoculated with *M. huakuii* 7653R.

Seeds were surface-sterilized with 3% (w/v) sodium hypochlorite for 30 min, and then rinsed with sterile distilled water for 3 times. The seedlings were grown in a sterile culture medium containing 6% (v/v) sucrose, 1% (w/v) agar, and 1% (w/v) casein hydrolysate in a greenhouse at 25 °C and 16 h light/8 h dark cycle at 18–22 °C. Six-d-old seedlings were inoculated with *M. huakuii* 7653R.

Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s protocol.
For gene expression analysis, RNA was isolated from infected roots at different dai (1, 3, 5, 7, 9, 12, 15, and 21 dai, respectively), roots with nodules removed (27 dai), nodules (27 dai), leaves (27 dai), leafstalks (27 dai), and uninfected roots (4, 6, and 33 d after sowing, respectively) to proceed to reverse transcription (RT)–PCR.

DNA was extracted from leaves by grinding in liquid N₂ with 10% (w/w) PVP, adding lysis buffer [1.5% (v/v) CTAB, 75 mM TRIS–HCl (pH 8), 15 mM EDTA (pH 8), 1.05 M NaCl, 4% (v/v) β-mercaptoethanol], preheating to 65 °C, heating for 30 min at 65 °C, and centrifuging for 5 min at 12 000 rpm. The supernatant was mixed with an equal volume of chloroform/isomyl alcohol (24:1, v/v) and blended gently upside down until milkiness appeared in the lower bottle-green liquid. After centrifugation, DNA was precipitated from the colourless supernatant with 0.1 vol. of 10 M ammonium acetate and 2 vols of 95% (v/v) alcohol for 10 min at room temperature, and centrifuged for 10 min at 12 000 rpm. The pellet was washed with 70% (v/v) ethanol, dried in air, and resuspended in sterile distilled deionized water.

**SMART cDNA synthesis**

cDNA was synthesized and amplified using the Clontech SMART PCR cDNA Synthesis Kit (Clontech, Palo Alto, CA, USA) with the following modification. Total RNA (1 µg) was reverse transcribed at 42 °C for 1 h in the presence of both cDNA synthesis (CDS) primer and SMART II oligonucleotide. Tricine-EDTA buffer (40 mM Tris–HCl (pH 8), 15 mM EDTA (pH 8), 1.05 M NaCl, 4% (v/v) β-mercaptoethanol), preheating to 65 °C, heating for 30 min at 65 °C, and centrifuging for 5 min at 12 000 rpm. The supernatant was mixed with an equal volume of chloroform/isomyl alcohol (24:1, v/v) and blended gently upside down until milkiness appeared in the lower bottle-green liquid. After centrifugation, DNA was precipitated from the colourless supernatant with 0.1 vol. of 10 M ammonium acetate and 2 vols of 95% (v/v) alcohol for 10 min at room temperature, and centrifuged for 10 min at 12 000 rpm. The pellet was washed with 70% (v/v) ethanol, dried in air, and resuspended in sterile distilled deionized water.

**Suppression subtractive hybridization and construction of a subtractive cDNA library**

The library was constructed using the Clontech PCR-Select cDNA Subtraction Kit. Briefly, 2 µg of poly(A) RNA were purified from total RNA of inoculated and non-inoculated roots at 21–26 dai with the PolyATract® mRNA Isolation System III (Promega, Madison, WI, USA) and then used as the template to synthesize two cDNA populations that were then digested with the RsaI restriction enzyme. The digested DNA fragments from each cDNA population were used as the tester cDNAs and driver cDNAs for forward and reverse subtraction. The cDNAs synthesized from inoculated roots were used as the tester in forward subtraction to clone genes specific for infected roots (data not shown), were constructed. Two subtractive cDNA libraries, one for infected roots and the other for uninfected control roots (data not shown), were constructed.

**Virtual northern blotting**

The cDNA fragments of candidate clones were amplified using T7 and SP6 promoter primers. The adaptors were then removed and used to prepare probes as described above. A 25 µl aliquot of SMART cDNAs was resolved in a 1.2% (w/v) agarose gel, denatured, and transferred to Hybond™-N+ nylon membrane. The membranes were hybridized as described above. A ubiquitin cDNA fragment (Madsen et al., 2003) was amplified and used as a loading control.

**RT–PCR**

Total RNA (1 µg) was treated with DNase I (RNase free, Takara) and used for cDNA synthesis and amplification with primers deduced from the SSH cDNA sequences (Table 1) in a final volume of 50 µl following the protocol of the supplier [One Step RNA PCR Kit (AMV), Takara]. First-strand cDNA was synthesized at 50 °C for 30 min, the enzyme was then inactivated for 2 min at 94 °C, and was followed by PCR cycles (denaturation 94 °C for 30 s; annealing, with the temperature depending on the primers for 30 s; elongation, 72 °C for 1 min) and a final 5 min extension step at 72 °C. Thirty PCR cycles were carried out. Aliquots (5 µl) of PCR products were analysed by electrophoresis in 2% agarose gels. As a constitutive control, the same procedure but with a total of 15 cycles was performed to amplify an 18S rRNA fragment. RT–PCR analysis was repeated at least twice with similar results.

**Full-length cDNA cloning by RACE**

Both 5’ and 3’ cDNA ends of the different fragments were obtained using the SMART™ RACE cDNA Amplification Kit (Clontech) and gene-specific primers (Table 2). The full-length cDNA or genomic DNA products could be assembled by overlapping 5’- and 3’-RACE fragments or generated by LD-PCR using primers designed from the extreme 5’ and 3’ ends of target cDNA and the 5’-RACE-Ready cDNA or genomic DNA as template, which were then cloned into pGEM-T vector (Promega).

**Sequence analysis**

The amino acids of target cDNA sequences were deduced and the sequences aligned using BioEdit software (Hall, 1999). Homology searches were performed through the BLAST programs (http://www.ncbi.nlm.nih.gov/, http://ca.expasy.org/), InterProScan (http://www.ebi.ac.uk/) (Quevillon et al., 2005) and Pfam databases (http://www.sanger.ac.uk/Software/) (Bateman et al., 2004) were used to identify conserved amino acid motifs. Signal peptides and putative cleavage sites were predicted by SignalP 3.0 Server (http://www.cbs.dtu.dk/services/) (Nielsen et al., 1998; Bendtsen et al., 2004). The theoretical pl (isoelectric point) and molecular weight were predicted with Compute pl/Mw tool (http://ca.expasy.org/tools/) (Gasteiger et al., 2005).

To produce probes, subtracted and unsubtracted cDNAs from infected or uninfected control roots were digested with Rsal to remove the adaptor sequences and were labelled with 32P using a Random Primer DNA Labeling kit (Takara, Dalian, China) following the manufacturer’s instructions. Membranes were prehybridized at 65 °C for 14 h in 5× SSC, 5× Denhardt’s, 0.5% (w/v) SDS and 100 µg ml⁻¹ salmon sperm DNA, hybridized overnight at 65 °C, then washed at 65 °C twice in 2× SSC/0.5% SDS, twice in 1× SSC/0.5% SDS, and twice in 0.1× SSC/0.5% SDS. The blots were exposed to X-ray film under an intensifying screen at –80 °C.

The differential fragments were selected by comparing the four identical membranes hybridized with four different cDNA probes. The SSH clones showing a more intense signal after hybridization to the cDNA probes from infected roots were sequenced for a similarity database search.
Characterization of the subtractive cDNA library and analysis of SSH cDNA sequences

Two subtracted cDNA libraries were generated by SSH. One was enriched with A. sinicus cDNA sequences that correspond to preferentially transcribed genes in the roots inoculated with M. huakuii and the other to those genes in non-inoculated roots (data not shown). The resulting SSH cDNA fragments were cloned, and a total of 400 clones were obtained.

Initial screening of the cDNAs to remove false positives was performed by dot blotting. The clones corresponding to genes with at least a 3-fold increase in transcript accumulation in infected roots compared with uninfected control roots were sequenced. Nineteen groups of different inserts were identified after annotation and redundancy analyses. BLAST searches indicated that 13 cDNA clones showed significant homology with Lb, and one clone, AsB2510, contained an open reading frame. Two fragments were identical with the A. sinicus asparagine synthetase gene and the other seven fragments were similar to the genes encoding ENOD2, nodulin-26, phosphoenolpyruvate carboxylase of Sesbania rostrata (very low similarities), putative cysteine proteinase of Capsicum annuum, diphosphonucleotide phosphatase 1 of Lupinus luteus, RNA helicase of Arabidopsis thaliana, and M. truncatula clone mth2-10i23, respectively. Moreover, 31 fragments did not show any significant match.

Confirmation of expression for selected clones

To verify the candidates as differentially expressed in the inoculated roots, virtual northern blotting was performed. The results showed (Fig. 1) that 12 clones examined were expressed only in the inoculated roots; the remaining clone, AsIB259, only displayed stronger signal intensities compared with control roots. The full-size cDNA sequences of the 13 clones were obtained by RACE and analysed by computer. Among them, AsC2411, AsA257, and AsD2410 showed two transcripts (Fig.1), but only one full-length cDNA was isolated.

Primary structure of the 13 genes under investigation

Sequence analysis of the 13 genes indicated that the 11 of them encode polypeptides similar to the CCPs (Frühling et al. 2001).

Table 1. Primer sequences used for RT-PCR and the expected sizes of RT-PCR products

<table>
<thead>
<tr>
<th>Gene</th>
<th>Size of products (bp)</th>
<th>Primer sequences for one-step RT-PCR (5’→3’)</th>
<th>Primer sequences for 5’ RACE</th>
<th>Primer sequences for 3’ RACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AsA244</td>
<td>310</td>
<td>GTCAAGTAACCTGTGGAAGAAG</td>
<td>GGTGTTGATTCTCAAGTGT</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
</tr>
<tr>
<td>AsA257</td>
<td>339</td>
<td>CAAAAGACATTGTGACAAATATC</td>
<td>GGGTAGAAGAGGAGGACACGAG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsC2411</td>
<td>247</td>
<td>CAAAAGTCTTAAAGCTTTCCTC</td>
<td>GCAACTTCTCAATGCTGTC</td>
<td>TTCCGCAGTGAATAAAAACCTGAGG</td>
</tr>
<tr>
<td>AsC259</td>
<td>202</td>
<td>CCAAAATGGGAGGAGAAGG</td>
<td>TTAACTCGTATGGATGATG</td>
<td>TTAAGACGAGATGAGATG</td>
</tr>
<tr>
<td>AsD2411</td>
<td>322</td>
<td>GGTTGTTGATTCTCAAGTGT</td>
<td>ACAAATGTTGATGGGCTGATGATGTA</td>
<td>TTAAGACGAGATGAGATG</td>
</tr>
<tr>
<td>AsD257</td>
<td>275</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsH2415</td>
<td>300</td>
<td>GTTGTTGATTCTCAAGTGT</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsI2412</td>
<td>287</td>
<td>CACATGAAAGGAGGAGGAGG</td>
<td>TTAAGACGAGATGAGATG</td>
<td></td>
</tr>
<tr>
<td>AsI257</td>
<td>403</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsJ2410</td>
<td>440</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsK247</td>
<td>224</td>
<td>GTTGTTGATTCTCAAGTGT</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsL246</td>
<td>368</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsM259</td>
<td>356</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>AsN257</td>
<td>399</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
<tr>
<td>18S rRNA</td>
<td>432</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>CACATAAGTTGAGCAAAATCTCAGTGC</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Gene-specific primer (GSP) sequences used for rapid amplification of cDNA ends (RACE)

<table>
<thead>
<tr>
<th>Gene</th>
<th>GSP sequences for 5’ RACE</th>
<th>GSP sequences for 3’ RACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AsA244</td>
<td>CATCGACACATCTTCTCAATGCCTG</td>
<td>CAACATAAGTTGAGCAAAATCTCAGTGC</td>
</tr>
<tr>
<td>AsA257</td>
<td>ATGCACCTCAGTCGGCAACAGGAGCAG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsC2411</td>
<td>TACACCCTTCAATCCCAAGGCTGCC</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsC259</td>
<td>TACACCCTTCAATCCCAAGGCTGCC</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsD2411</td>
<td>CCACACATATTGGTTCAACCCGGGCAAGCCA</td>
<td>ACAAATGTTGATGGGCTGATGATGTA</td>
</tr>
<tr>
<td>AsD257</td>
<td>CGCATAAACCGGGAGGAGCAATAAAG</td>
<td>TTAAGACGAGATGAGATG</td>
</tr>
<tr>
<td>AsH2415</td>
<td>CATATCGATCAAGTCTTCAATCCCAAGGCTGCC</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsI2412</td>
<td>CAGTCTGACATTTCTTCAATGCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsI257</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsJ2410</td>
<td>TTTGTGGATATCTTCTTCAATGCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsK247</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsL246</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsM259</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>AsN257</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
<td>TCCCTGAAAAATGGGCAGACGTCCTG</td>
</tr>
<tr>
<td>18S rRNA</td>
<td>432</td>
<td>GATCTGACATTTCTTCAATGCCTG</td>
</tr>
</tbody>
</table>

Results

Characterization of the subtractive cDNA library and analysis of SSH cDNA sequences

Two subtracted cDNA libraries were generated by SSH. One was enriched with A. sinicus cDNA sequences that correspond to preferentially transcribed genes in the roots inoculated with M. huakuii and the other to those genes in non-inoculated roots (data not shown). The resulting SSH cDNA fragments were cloned, and a total of ~400 clones were obtained.

Initial screening of the cDNAs to remove false positives was performed by dot blotting. The clones corresponding to genes with at least a 3-fold increase in transcript accumulation in infected roots compared with uninfected control roots were sequenced. Nineteen groups of different inserts were identified after annotation and redundancy analyses. BLAST searches indicated that 13 cDNA clones showed significant homology with Lb, and one clone, AsB2510, contained an open reading frame. Two fragments were identical with the A. sinicus asparagine synthetase gene and the other seven fragments were similar to the genes encoding ENOD2, nodulin-26, phosphoenolpyruvate carboxylase of Sesbania rostrata (very low similarities), putative cysteine proteinase of Capsicum annuum, diphosphonucleotide phosphatase 1 of Lupinus luteus, RNA helicase of Arabidopsis thaliana, and M. truncatula clone mth2-10i23, respectively. Moreover, 31 fragments did not show any significant match.

Confirmation of expression for selected clones

To verify the candidates as differentially expressed in the inoculated roots, virtual northern blotting was performed. The results showed (Fig. 1) that 12 clones examined were expressed only in the inoculated roots; the remaining clone, AsB259, only displayed stronger signal intensities compared with control roots. The full-size cDNA sequences of the 13 clones were obtained by RACE and analysed by computer. Among them, AsC2411, AsA257, and AsD2410 showed two transcripts (Fig.1), but only one full-length cDNA was isolated.

Primary structure of the 13 genes under investigation

Sequence analysis of the 13 genes indicated that the 11 of them encode polypeptides similar to the CCPs (Frühling et al. 2001).
et al., 2000; Mergaert et al., 2003), the similarities including a small size (58–126 amino acids), the presence of a conserved signal peptide at the N-terminus, and conserved Cys-X4-Asp-Cys and Cys-X4-5-Cys elements in their C-terminal halves. In addition, a transmembrane region was also revealed at the N-terminal domain by InterProScan.

Despite the sequences of their signal peptide domain being highly conserved, the alignment of the 11 deduced polypeptide sequences displayed relatively low homologies at the amino acid level (Fig. 2). A hydrophobic N-terminal domain of 25–30 amino acids of the 11 sequences was predicted as a signal peptide with high probability, and a cleavage site for removal of the signal peptide was predicted for nine of them, the exceptions being AsA257 and AsF259 for which a signal anchor was predicted with the SignalP program. According to the cysteine motif, the 11 genes could be divided into three major groups: the first group, comprising AsA244, AsA257, AsC2411, AsF259, AsG2411, AsIC258, AsIIA255, and AsIIC2512, contained two pairs of cysteines, whereas the second group involving AsA244, AsG2411, and AsD255 had six conserved cysteines. AsD2410 with eight cysteines constituted the third group. The amino acids around the cysteine residues resembled the family NCR (nodule-specific cysteine rich) (Mergaert et al., 2003). However, a few deviations and new features in the structure of both cysteine clusters were found (Fig. 2).

Two types of changes were found: (i) a charged residue (lysine or glutamate) more predominant than proline following C4; and (ii) glycine and a proximate aromatic residue (tryptophan or tyrosine) between C4 (or C5) and C6 were conserved even more than proline. Two new features were observed: (i) aspartate commonly appeared at the third amino acid after C3; and (ii) a hydrophobic residue commonly appeared as the first amino acid after C6 and a conserved acidic amino acid (aspartate or glutamate) as the second (Fig. 2).

Motif searches using Pfam and InterProScan revealed that the eight shorter sequences (AsA244, AsA257, AsC2411, AsF259, AsG2411, AsIC258, AsIIA255, and AsIIC2512, 59–70 amino acids) showed significant matches with several late nodulin sequences which are homologous to the Pisum sativum (garden pea) ENOD3 protein (Table 3; Scheres et al., 1990). AsG257 and AsD255 were identified to contain a scorpion toxin-like domain (0.0018 and 0.00027, respectively) at the C-terminus, and AsG257 had a trusted match (domains scoring higher than the gathering threshold) with the defensin 2 domain (1.2e-05). No significant matches with AsD2410 were detected (Table 3).

The genomic DNA sequences of AsD255 and AsA257 were isolated. When compared with the full-length cDNA sequences, two introns were identified in AsD255. One with 125 nucleotides was displayed between the first and second nucleotide of the 61st codon (Leu61; Fig. 3), which separated the first exon from about three-fifths of the full-length polypeptide sequence of AsD255. Another 91 nucleotide intron was beyond the open reading frame of AsD255 and between the second and third nucleotide of one stop codon (Fig. 3). Unexpectedly, no intron was found in the genomic DNA sequence of AsA257.

The other two cDNAs, AsE246 and AsIB259, were revealed to contain a AAI_plantLTP and AAI_LTSS domain. A signal peptide and the cleavage site were predicted for them. A transmembrane region was also discovered at the N-terminal domain. Although they were
predicted to come from the same protease inhibitor/seed storage/LTP family, the amino acid sequences deduced from the two cDNAs shared only 19% identities. The BLASTP program revealed that AsE246 was homologous to the non-specific LTP from *Lycopersicon esculentum* with 61% similarity (42% identity), LTP 1 from *Euphorbia lagascae* with 58% similarity (43% identity), and AsIB259 homologous to protease inhibitor/seed storage/LTP family protein from *Arabidopsis thaliana* with 48% similarity (35% identity). Sequence alignment showed that the predicted AsE246 protein had significant similarity with LTPs from various plants (Fig. 4A), whereas AsIB259 more closely resembled several unknown proteins or LTPs of *A. thaliana* or *Oryza sativa* (Fig. 4B). In contrast, AsE246 and AsIB259 displayed lower homology to the LTPs from legumes at the amino acid level (Fig. 4C).

The theoretical molecular weight and isoelectric point (pI) of the putative proteins from the 13 genes are shown in Table 4.

Expression analysis of the genes identified
In order to analyse gene expression profiles during nodule development, RT–PCR was performed (Figs 5, 6). A new

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession no.</th>
<th>Length of putative polypeptide (no. of amino acids)</th>
<th>Best homology (BLASTP, NCBI) and E value</th>
<th>Domain analysis with Pfam (trusted matches) and E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AsA244</td>
<td>DQ199634</td>
<td>70</td>
<td>Hypothetical protein, <em>Galega orientalis</em> CAB51773, 0.002</td>
<td>Nodulin late, 1e-08</td>
</tr>
<tr>
<td>AsA257</td>
<td>DQ199641</td>
<td>68</td>
<td>ENOD3, <em>Pisum sativum</em> AAB23537, 0.18</td>
<td>Nodulin late, 9e-14</td>
</tr>
<tr>
<td>AsC2411</td>
<td>DQ199638</td>
<td>60</td>
<td>ENOD3, <em>Pisum sativum</em> AAB23537, 0.23</td>
<td>Nodulin late, 9e-14</td>
</tr>
<tr>
<td>AsF259</td>
<td>DQ199639</td>
<td>64</td>
<td>Late nodulin, <em>Vicia faba</em> CAB96475, 0.036</td>
<td>Nodulin late, 7.2e-11</td>
</tr>
<tr>
<td>AsG2411</td>
<td>DQ199636</td>
<td>63</td>
<td>Hypothetical protein, <em>Galega orientalis</em> CAB51773, 1e-04</td>
<td>Nodulin late, 1.4e-13</td>
</tr>
<tr>
<td>AsG257</td>
<td>DQ199643</td>
<td>70</td>
<td>ENOD3, <em>Pisum sativum</em> AAB23537, 0.02</td>
<td>Nodulin late, 6.7e-14</td>
</tr>
<tr>
<td>AsD255</td>
<td>DQ199650</td>
<td>101</td>
<td>No hits below 1e</td>
<td>Defensin 2, 1.2e-05</td>
</tr>
<tr>
<td>AsD2410</td>
<td>DQ199646</td>
<td>126</td>
<td>No hits below 1e</td>
<td>Tryp_alpha_amyl, 1.9e-11</td>
</tr>
<tr>
<td>AsG257</td>
<td>DQ199643</td>
<td>90</td>
<td>No hits below 1e</td>
<td>Defensin 2, 1.2e-05</td>
</tr>
<tr>
<td>AsE246</td>
<td>DQ199648</td>
<td>132</td>
<td>NLTTP, <em>Lycopersicon esculentum</em> CAJ19705, 0.002</td>
<td>Nodulin late, 7.3e-14</td>
</tr>
<tr>
<td>AsG259</td>
<td>DQ199649</td>
<td>148</td>
<td>Unknown protein <em>Arabidopsis thaliana</em> AAL66904, 0.1e-11</td>
<td>Nodulin late, 7.3e-14</td>
</tr>
<tr>
<td>AsD2510</td>
<td>DQ199647</td>
<td>148</td>
<td>Leghaemoglobin, <em>Sesbania rostrata</em> CAA31859, 0e-54</td>
<td>Nodulin late, 7.3e-14</td>
</tr>
</tbody>
</table>

Fig. 2. Alignment of the deduced polypeptide sequences of *Astragalus sinicus* late nodulins AsA257, AsC2411, AsF259, AsG2411, AsIC258, AsIA255, AsIC2512, AsA244, AsG257, AsD255, and AsD2410, the early pea nodulin PeENOD3 (AAB23537), and *Vicia faba* late nodulin Vf-CCP1 (CAB96472). Identical residues conserved in at least 50% of the sequences are in black, whereas similar amino acids are on a grey background. The signal peptides are underlined. The conserved cysteines are numbered as described in the text.
set of infected and uninfected control plants were grown as described for SSH. RNA was extracted from the infected roots, and from control roots and other organs. All RNA preparations were checked for DNA contamination by PCR using 18S rDNA primers (data not shown). Comparative RNA accumulation levels of the 13 genes are shown in Figs 5 and 6. To distinguish between induction of the genes caused by the infection process and by development of young roots, RNA samples corresponding to 4-d-old and 6-d-old uninfected control roots were included in the RT–PCR analysis. Except for AsC2411 and AsIB259, no visible products were obtained after amplification, indicating that the gene induction was infection specific (Fig. 5). The AsC2411 gene showed only a basal expression in all control roots which was similar to that detected in young infected roots, and it was strongly induced during nodule development (Figs 5, 6). The AsIB259 gene was not expressed in the young roots, but was expressed in mature 33-d-old control roots as well as nodules and infected roots (Figs 5, 6). As a symbiotic marker characteristic in functional nodules, the A. sinicus leghaemoglobin gene AsB2510 was used, and its expression started ~5 dai. With the exception of AsIB259, which was similar to AsB2510, the remaining 12 genes were detected 7–9 dai (Fig. 5). Their transcriptional levels were very high in root nodules, and some of them could also be detected in infected roots from which the nodules had been removed at decreased levels. No amplification products could be detected in leaves and leafstalks.

Discussion

Fifty-three gene clones expressed differentially in the inoculated roots of Chinese milk vetch compared with uninfected control roots were isolated by the SSH procedure. Of these clones, the full-length cDNA sequence of 13 genes were obtained by RACE and analysed by computer. Although no significant identity with any published sequences was displayed over the entire length of the predicted polypeptides, 11 of the 13 genes did share certain structural features with some previously reported nodulin genes, mainly including a relatively well conserved signal peptide at the N-terminus and a conserved cysteine motif at the C-terminus. In addition, RT–PCR analysis showed that they were expressed 2–4 d later than the leghaemoglobin gene AsB2510 of A. sinicus during nodule development and so were classified as late nodulin genes according to Nap and Bisseling (1990).

Mergaert et al. (2003) suggested that NCR polypeptide genes had introns presenting at a conserved position and that they were all nodule specific. We also verified a 125 nucleotide intron in AsD255, but no intron was detected in AsA257. Kato et al. (2002) characterized five CCP genes with reduced expression in ineffective nodules in G. orientalis (Kaijalaainen et al., 2002). Fifty-three gene clones expressed differentially in the inoculated roots of Chinese milk vetch compared with uninfected control roots were isolated by the SSH procedure. Of these clones, the full-length cDNA sequence of 13 genes were obtained by RACE and analysed by computer. Although no significant identity with any published sequences was displayed over the entire length of the predicted polypeptides, 11 of the 13 genes did share certain structural features with some previously reported nodulin genes, mainly including a relatively well conserved signal peptide at the N-terminus and a conserved cysteine motif at the C-terminus. In addition, RT–PCR analysis showed that they were expressed 2–4 d later than the leghaemoglobin gene AsB2510 of A. sinicus during nodule development and so were classified as late nodulin genes according to Nap and Bisseling (1990).

Mergaert et al. (2003) suggested that NCR polypeptide genes had introns presenting at a conserved position and that they were all nodule specific. We also verified a 125 nucleotide intron in AsD255, but no intron was detected in AsA257. Interestingly, AsA257 and AsF259 were predicted to have a signal anchor. The signal anchor was an uncleaved signal peptide and the resulting protein was named as a type II membrane protein (Nielsen and Krogh, 1998). In contrast to previously reported CCPs and the other 10 genes encoding these proteins, the transcript of AsC2411 was detected in uninfected control roots by RT–PCR which is more sensitive than northern analysis (Figs 5, 6).

Nodule-specific CCP homologues have been identified in V. faba (Frühling et al., 2000) and G. orientalis (Kaijalaainen et al., 2002). Kato et al. (2002) characterized five CCP genes with reduced expression in ineffective nodules in G. orientalis (Kaijalaainen et al., 2002). Fifty-three gene clones expressed differentially in the inoculated roots of Chinese milk vetch compared with uninfected control roots were isolated by the SSH procedure. Of these clones, the full-length cDNA sequence of 13 genes were obtained by RACE and analysed by computer. Although no significant identity with any published sequences was displayed over the entire length of the predicted polypeptides, 11 of the 13 genes did share certain structural features with some previously reported nodulin genes, mainly including a relatively well conserved signal peptide at the N-terminus and a conserved cysteine motif at the C-terminus. In addition, RT–PCR analysis showed that they were expressed 2–4 d later than the leghaemoglobin gene AsB2510 of A. sinicus during nodule development and so were classified as late nodulin genes according to Nap and Bisseling (1990).

Mergaert et al. (2003) suggested that NCR polypeptide genes had introns presenting at a conserved position and that they were all nodule specific. We also verified a 125 nucleotide intron in AsD255, but no intron was detected in AsA257. Interestingly, AsA257 and AsF259 were predicted to have a signal anchor. The signal anchor was an uncleaved signal peptide and the resulting protein was named as a type II membrane protein (Nielsen and Krogh, 1998). In contrast to previously reported CCPs and the other 10 genes encoding these proteins, the transcript of AsC2411 was detected in uninfected control roots by RT–PCR which is more sensitive than northern analysis (Figs 5, 6).

Nodule-specific CCP homologues have been identified in V. faba (Frühling et al., 2000) and G. orientalis (Kaijalaainen et al., 2002). Kato et al. (2002) characterized five CCP genes with reduced expression in ineffective nodules in G. orientalis (Kaijalaainen et al., 2002). Fifty-three gene clones expressed differentially in the inoculated roots of Chinese milk vetch compared with uninfected control roots were isolated by the SSH procedure. Of these clones, the full-length cDNA sequence of 13 genes were obtained by RACE and analysed by computer. Although no significant identity with any published sequences was displayed over the entire length of the predicted polypeptides, 11 of the 13 genes did share certain structural features with some previously reported nodulin genes, mainly including a relatively well conserved signal peptide at the N-terminus and a conserved cysteine motif at the C-terminus. In addition, RT–PCR analysis showed that they were expressed 2–4 d later than the leghaemoglobin gene AsB2510 of A. sinicus during nodule development and so were classified as late nodulin genes according to Nap and Bisseling (1990).

Mergaert et al. (2003) suggested that NCR polypeptide genes had introns presenting at a conserved position and that they were all nodule specific. We also verified a 125 nucleotide intron in AsD255, but no intron was detected in AsA257. Interestingly, AsA257 and AsF259 were predicted to have a signal anchor. The signal anchor was an uncleaved signal peptide and the resulting protein was named as a type II membrane protein (Nielsen and Krogh, 1998). In contrast to previously reported CCPs and the other 10 genes encoding these proteins, the transcript of AsC2411 was detected in uninfected control roots by RT–PCR which is more sensitive than northern analysis (Figs 5, 6).

Nodule-specific CCP homologues have been identified in V. faba (Frühling et al., 2000) and G. orientalis (Kaijalaainen et al., 2002). Kato et al. (2002) characterized five CCP genes with reduced expression in ineffective nodules in G. orientalis (Kaijalaainen et al., 2002). Fifty-three gene clones expressed differentially in the inoculated roots of Chinese milk vetch compared with uninfected control roots were isolated by the SSH procedure. Of these clones, the full-length cDNA sequence of 13 genes were obtained by RACE and analysed by computer. Although no significant identity with any published sequences was displayed over the entire length of the predicted polypeptides, 11 of the 13 genes did share certain structural features with some previously reported nodulin genes, mainly including a relatively well conserved signal peptide at the N-terminus and a conserved cysteine motif at the C-terminus. In addition, RT–PCR analysis showed that they were expressed 2–4 d later than the leghaemoglobin gene AsB2510 of A. sinicus during nodule development and so were classified as late nodulin genes according to Nap and Bisseling (1990).

Mergaert et al. (2003) suggested that NCR polypeptide genes had introns presenting at a conserved position and that they were all nodule specific. We also verified a 125 nucleotide intron in AsD255, but no intron was detected in AsA257. Interestingly, AsA257 and AsF259 were predicted to have a signal anchor. The signal anchor was an uncleaved signal peptide and the resulting protein was named as a type II membrane protein (Nielsen and Krogh, 1998). In contrast to previously reported CCPs and the other 10 genes encoding these proteins, the transcript of AsC2411 was detected in uninfected control roots by RT–PCR which is more sensitive than northern analysis (Figs 5, 6).

Nodule-specific CCP homologues have been identified in V. faba (Frühling et al., 2000) and G. orientalis (Kaijalaainen et al., 2002). Kato et al. (2002) characterized five CCP genes with reduced expression in ineffective nodules in G. orientalis (Kaijalaainen et al., 2002). Fifty-three gene clones expressed differentially in the inoculated roots of Chinese milk vetch compared with uninfected control roots were isolated by the SSH procedure. Of these clones, the full-length cDNA sequence of 13 genes were obtained by RACE and analysed by computer. Although no significant identity with any published sequences was displayed over the entire length of the predicted polypeptides, 11 of the 13 genes did share certain structural features with some previously reported nodulin genes, mainly including a relatively well conserved signal peptide at the N-terminus and a conserved cysteine motif at the C-terminus. In addition, RT–PCR analysis showed that they were expressed 2–4 d later than the leghaemoglobin gene AsB2510 of A. sinicus during nodule development and so were classified as late nodulin genes according to Nap and Bisseling (1990).
nODULES ON P. SATIVUM AND THOUGHT THAT THEIR SUCCESSIVE
EXPRESSION DURING NODULE DEVELOPMENT WAS ASSOCIATED
WITH NITROGEN-FIXING ACTIVITY. CROCKARD ET AL. (2002) ISO-
LATED AN EARLY NODULIN HOMOLOGOUS TO CCPS FROM WHITE
CLOVER ROOT AND SUGGESTED THAT dd23b WAS RELATED TO
METAL ION TRANSPORT IN THE PERIBACTEROID MEMBRANE AND TO
PLANT CELL INFECTION. FROM M. TRUNCATULA, 114 NODULE-
SPECIFIC TCs SHOWING SIMILARITY TO CCPs WERE REPORTED
BY FEDOROVA ET AL. (2002). SUBSEQUENTLY, AN EXTREMELY
LARGE AND NODULE-SPECIFIC CCP FAMILY COMPOSED OF
>300 GENES WAS IDENTIFIED IN M. TRUNCATULA BY MERGAERT
PUTATIONAL APPROACHES. FURTHER, GRAHAM ET AL. POINTED
OUT THAT CCPs WERE NOT SPECIFIC TO GALEGOID LEGUMES AND
FUNCTIONED IN IDENTERMINATE NODULE FORMATION AS HYPOTHESIZED BY MERGAERT ET AL. (2003) BECAUSE HOMOLOGOUS
SEQUENCES WERE DETECTED IN THE SOYBEAN GENOME.

DUE TO MOTIFS SIMILAR TO PLANT DEFENSINS, THE FUNCTION OF
CCPs AS ANTIMICROBIAL DEFENSINS WAS SUGGESTED (MERGAERT
EFFECTORS OF INNATE IMMUNITY, DEFENSINS ARE WIDELY DISTRIB-
UTED IN THE ANIMAL AND PLANT KINGDOMS AND THOUGHT TO BE
MEMBERS OF SMALL (15–40 MEMBERS) GENE FAMILIES (BOMAN,
2003). PLANT DEFENSINS ARE SMALL, GENERALLY BASIC PROTEINS
OF 45–54 AMINO ACIDS AND THEIR TERTIARY STRUCTURES SHARE
A COMMON \( \alpha/\beta \) MOTIF STABILIZED BY EIGHT CYSTEINE RESIDUES
AND COMPOSED OF THREE ANTIPARALLEL \( \beta \)-SHEETS AND ONE \( \alpha \)-HELIX
(ALMEIDA ET AL., 2002). THIS KIND OF STRUCTURAL PATTERN
WAS ALSO FOUND IN INSECT DEFENSINS AND SOME SCORPION
neurotoxins (Kobayashi et al., 1991; Cornet et al., 1995). The plant sequences containing this motif have different functions ranging from storage, protection, enzyme inhibition and lipid transfer, to cell wall structure (José-Estanyol et al., 2004). Since the plant defensins were first isolated from wheat and barley grains (Colilla et al., 1990; Mendez et al., 1990), more plant defensins have been isolated from a variety of tissues including leaves (Terras et al., 1995), flowers (Park et al., 2002), and seeds (Almeida et al., 2000; Chen et al., 2002) of different plants. Recently, Hanks et al. (2005) reported 16 putative cysteine-rich defensins by using a cloned seed defensin from M. sativa as the BLAST query for the Institute for Genomic Research's (TIGR) M. truncatula gene index (MtGI version 7). By searching in the almost completed A. thaliana genome using motif models, >300 defensin-like genes were identified in Arabidopsis (Silverstein et al., 2005).

Indeed, motif analysis has led to the identification of AsG257 as attributable to the defensin 2 family, which is a member of the scorpion toxin-like knottin superfamily and contains six conserved cysteines all involved in intrachain disulphide bonds. Although AsA244, AsD255, and AsG257 have similar conserved cysteine elements, the Cys-X5-Cys-X3-Cys-X9-Cys-X5-Cys-X1-Cys pattern of AsD255 and AsG257 was more coincident with the cysteine motif of the scorpion toxin-like domain than the Cys-X5-Cys-X3-Cys-X7-Cys-X4-Cys-X7-Cys motif of AsA244. For example, the arthropod defensin family including insect and scorpion cysteine-rich antibacterial peptides have a Cys-Xn-Cys-X3-Cys-X9-Cys-X5-Cys-X1-Cys motif (Bulet et al., 1992; Yamada and Natori, 1993). In addition, the charge of the 11 mature proteins differed from positive to negative to nearly neutral (Table 4). The results in Table 4 showed that the processed polypeptides AsA244, AsD2410, and AsG257 were basic and AsD255 was almost neutral. The remaining seven were all acidic. The different amino acids around the cysteine residues and different charge situation could explain why the other CCP-like proteins with the exception of AsG257 had no significant hit with the defensin family.

Like AsD2410, AsE246 and AsIB259 also contain eight cysteines. However, the eight-cysteine pattern of AsE246 and AsIB259 is distinct from that of AsD2410. The last six cysteines of AsD2410 constitute the same model as AsD255 and AsG257 (Fig. 2). The polypeptides of AsE246 and AsIB259 show a similar pattern in a region containing eight cysteine residues to the 8CM (eight-cysteine motif) family proteins. This protein has a signal peptide at the N-terminus and a conserved eight cysteine motif formed by <100 amino acid residues near the C-terminal end, with the consecutive third and fourth cysteines in the polypeptide chain and the fifth and sixth cysteines separated by only one residue (José-Estanyol et al., 2004). The 8CM family proteins include glutamine-rich 2S-albumins, cell wall proline-rich structural proteins termed hybrid proline-rich proteins (HyPRPs), and LTPs with the same 8CM domain but with a hydrophilic character (José-Estanyol et al., 2004).

Two LTP-like genes, AsE246 and AsIB259, were identified in this study. Apart from sequence homology with LTPs from different plants (Fig. 4A), other characteristics of AsE246 such as an evaluated molecular weight of 12.1 kDa, an isoelectric point of 8.66, and a signal peptide with a hydrophobic region at the N-terminal end are similar to most LTPs (Kader, 1996). In contrast, the molecular weight and isoelectric point of AsIB259 are higher than that of AsE246 (Table 4). There are four additional amino acid residues between the first and second cysteine of AsIB259 that is more in agreement with the several LTP-like proteins of A. thaliana or O. sativa (Fig. 4B). Expression analysis showed that the two LTP-like genes had a different expression profile. The AsE246 gene was expressed strongly in the inoculated roots at 9 dai and the transcript of AsIB259 was detected at 5 dai and accumulated abundantly at 7 dai (Fig. 5). Moreover, the expression of AsE246 was restricted to nodules, which indicated that the unique LTP-like gene was specific to nodule development. The transcription of AsIB259 detected in mature roots but not in young roots indicated that the LTP-like gene could be related to root development as well as to the infection process.

LTP-like genes have been isolated from several legumes (Choi et al., 1996; Liu et al., 2003; Carvalho et al., 2004). Krause et al. (1994) isolated a gene encoding the LTP-like protein from Vigna unguiculata root, AKCS9, which was expressed in all tissues tested except nodules. In recent studies, no LTP transcripts were found in nodules of P. vulgaris or in nodules of V. unguiculata (Verdoy et al., 2004). Another LTP-like protein MtN5 was reported to be one of the early markers expressed before infection during root nodule development of M. truncata (Gamars et al., 1996; El Yahyaoui et al., 2004). Compared with the LTPs from non-leguminous plants, AsE246 and AsIB259 show very low similarities to AKCS9 and MtN5 (Fig. 4C). AsE246 and AsIB259 should be the first nodule-specific or enhanced examples of LTP genes that are induced during the late stages of interaction between leguminous plants and rhizobia.

LTPs have been shown to be involved in plant biotic and abiotic stresses (Garcia-Olmedo et al., 1995) and transfer of lipids from one membrane to another in vitro (Kader, 1996). They were found to be secreted and located in the cell wall and could exert different functions such as furnishing cutin monomers to the cuticle (Sterk et al., 1991; Thoma et al., 1993), involving the permeabilization of membranes during the antifungal response (Cammue et al., 1995; Kader, 1996), and binding to specific membrane lipoproteins or to eliciting membrane receptor sites (Blein et al., 2002). Some members of the LTP family are known food allergens or have a relationship to asthma (Sanchez-Monge et al., 1992; Pastorello et al., 2001).
Why are so many antimicrobe-like proteins such as CCPs including AsA244, AsA257, AsC2411, AsD2410, AsD255, AsF259, AsG2411, AsIC258, AsIIA255, and AsIIIC2512, defensin-like protein AsG257 and LTP-like proteins AsE246, AsIB259, involved in root nodules? Apparently, the encoded proteins are needed in high quantities because they are expressed abundantly in nodules. Their different expression patterns suggest that they are differentially regulated. During the early steps of nodule formation, typical defence responses such as the production of phytoalexins and pathogenesis-related proteins are suppressed in the plant (Mithofer, 2002; Mitra and Long, 2004). Graham et al. (2004) suggested that the CCP genes could be induced as a secondary defence to avoid opportunistic infections by other soil micro-organisms and protect the nodule from pathogenic organisms during nodule formation when typical defence responses are shut down. Their extreme sequence variations could target specific pathogens without harming beneficial rhizobia while providing broad-spectrum antimicrobial activity. However, the plant defensins are generally basic proteins (Thomma et al., 2002) whereas eight of the 11 CCPs are acid to neutral. De Samblanx et al. (1997) revealed that the activity of Rs-AFP2 against filamentous fungi could be increased by the substitution of two neutral residues by arginines, and the replacement of a basic lysine residue by a neutral glutamine residue decreased its antifungal activity. It has been demonstrated that there might be no correlation between the three-dimensional structure of plant defensins and their action mechanisms that could be related to surface distribution of amino acid residues (Almeida et al., 2002; Melo et al., 2002). Considering that the peptides have been shown to bear signalling activity in plants (Lindsey et al., 2002), some of the antimicrobe-like polypeptides may not be acting as antimicrobial factors but as signals to mediate nodule formation and development. In fact, Maldonado et al. (2002) have identified a putative LTP protein that could advance long distance signalling by interacting with a lipid-derived molecule in Arabidopsis. The sequence similarity between the nodulation-related proteins and the proteins associated with plant defence supports the theory of a common origin of plant defence and symbiotic responses. Originally, these proteins might function as the common signalling components involved in plants’ perception of soil microorganisms and then could functionally differentiate whether to ‘accept’ or ‘refuse’ reaction during the involvement.

Acknowledgements

This work was supported by grants from the National 973 Key Fundamental Research Program of China (Grant No. 01CB108901) and the National Natural Science Foundation of China.

References


