



RESEARCH PAPER

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

* To whom correspondence should be addressed. E-mail: zhoujunchu@mail.hzau.edu.cn

Abbreviations: CCP, cysteine cluster protein; dai, days after inoculation; LD-PCR, long-distance polymerase chain reaction; LTP, lipid transfer protein; NCR, nodule-specific cysteine-rich; RT–PCR, reverse transcription–polymerase chain reaction; SSH, suppressive subtractive hybridization; TC, tentative consensus sequence.

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 cysteine 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 (TCs) or singletons belonging to a gene family named *NCR* (nodule-specific cysteine-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 cysteine 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 *NCR* genes with the genes coding for a potentially secreted calmodulin-like protein and for a signal peptide peptidase. Besides galeoid 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* (Kaijalainen *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.

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-cum-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, *AsNODc22*, 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 cysteine proteinase.

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 domain. The remaining two genes, *AsE246* and *AsIB259*, code for lipid transfer protein (LTP)-like proteins. Possible functions of these genes in nodule formation are discussed.

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-day-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 control plant roots at 21–26 days after inoculation (dai) by using TRIzol[®] 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% (w/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/isoamyl 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 μ g) 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 μ l) was added to 10 μ l of the first-strand reaction product and incubated at 72 °C for 7 min. Then 1 μ l of the diluted cDNA was added to a 100 μ l reaction mixture for long-distance (LD) PCR. The LD-PCR was conducted at 95 °C for 15 s, 65 °C for 30 s, and 68 °C for 6 min on a PTC-100™ Peltier Thermal Cycler for 17 cycles. This PCR product was used for virtual northern blotting.

Suppression subtractive hybridization and construction of a subtractive cDNA library

The library was constructed using the Clontech PCR-Select cDNA Subtraction Kit. Briefly, 2 μ g of poly(A) RNA were purified from total RNA of inoculated and non-inoculated roots at 21–26 dai with the PolyATtract® 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 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 the infected roots, and those from non-inoculated roots were used as the tester in reverse subtraction.

Two subtractive cDNA libraries, one for infected roots and the other for uninfected control roots (data not shown), were constructed.

Differential screening by dot blotting

Inserts of the SSH library clones were amplified from 1 μ l of plasmids using T7 and SP6 promoter primers. An 8 μ l aliquot of amplification products was denatured with 0.4 N NaOH (freshly made) and 10 mM EDTA (pH 8.2). A 2 μ l aliquot of each mixture was blotted onto a Hybond™-N⁺ nylon membrane (Amersham Pharmacia Biotech Limited, Little Chalfont, Buckinghamshire, UK) and four identical blots were made. The blotted membranes were washed in 2× SSC and the DNA was cross-linked to the membrane by baking the blots for 2 h at 80 °C in an oven. Four sets of membranes were prepared for hybridization: one was hybridized with subtracted cDNA probes made from infected roots, the second with unsubtracted cDNA probes from infected roots, and the other two with subtracted and unsubtracted cDNA probes from uninfected control roots, respectively.

To produce probes, subtracted and unsubtracted cDNAs from infected or uninfected control roots were digested with *RsaI* 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 μ g 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.

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.

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 μ l 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 μ g) 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 μ l 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 μ l) 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 searched 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 pI (isoelectric point) and molecular weight were predicted with Compute pI/Mw tool (<http://ca.expasy.org/tools/>) (Gasteiger *et al.*, 2005).

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

Gene	Size of products (bp)	Primer sequences for one-step RT-PCR (5'–3')
<i>AsA244</i>	310	GTCAAGTAACTGTGGAAAG
<i>AsA257</i>	339	CAAAGACATTGTTACAAATATC
<i>AsC2411</i>	247	CAAAGTCATTGTTACAAGTCT
<i>AsF259</i>	202	CCAAAATTGGGAGAGATATG
<i>AsG2411</i>	322	GGGTGGGAATATGCCTTTA
<i>AsIC258</i>	275	CATTATTGCAAGTCCCATAAA
<i>AsIIA255</i>	300	GTGGTAATTATGAAGCTTGT
<i>AsIIC2512</i>	287	CACGTAAGGAAAGCCCAA
<i>AsD255</i>	493	CTGCAATGATTCTCTTTT
<i>AsD2410</i>	440	GGATGAAGTTTTGTAGTAATC
<i>AsG257</i>	224	GTGTCAAGTGTATTTCTACG
<i>AsE246</i>	368	GCAATATAGAAATCCCCAAG
<i>AsIB259</i>	356	GTGATGAACATGAAGTTTAAG
<i>AsB2510</i>	399	GTTTCACTGAGGCACAA
<i>18S rRNA</i>	432	GACAGACTGAGAGCTCTT

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

Gene	GSP sequences for 5' RACE	GSP sequences for 3' RACE
<i>AsA244</i>	CATCGACACATTCTTTCAATGCACTTCGGTTT	CAACTAACGTTGACGCAATTCGTGTGCA
<i>AsA257</i>	ATGCACCTCACTGTCCAACCAAGGAGCAG	TCCTGAAATTGGGGACACTGCTCCTGGT
<i>AsC2411</i>	TATGCACTTCAATTCCCAACCAAGGTCC	TCCGACTGTAAAAAACTTTGGACGCCG
<i>AsF259</i>	TACACTTCGATTTCACCAACAGAAGGTGCTG	TTAACGCTATTTCGATGTAAAGATGATGCCG
<i>AsG2411</i>	CCACACATATTGTTCCATAACCAGCGAAGCCA	AACAAATGTTGATGGGGCTGATGAATGTATAA
<i>AsIC258</i>	CGCATAACCAGGGGATGGCACATAAAGA	TCAATGTGAAACGGATGAAGACTGTAAGGG
<i>AsIIA255</i>	CATAATCGATCAATGCACTTCGCTTCCCAA	ACGCTATGCCATGTGAAACAGATGACGACT
<i>AsIIC2512</i>	CAGTCGACATCATTTATACATGGCATACTGG	TCCCTGTTTCTTATGGCAATTAATGTTAACGC
<i>AsD255</i>	GGTTCAGCAGAGACACTTCCTTTTCGG	TAGAGGAAACGGGAAAAAAGGGGGAAA
<i>AsD2410</i>	CCACCGTAGAACGCAGTAACATAATCCCC	GGGGATTAGTTACTGCGTTCTACGGTGG
<i>AsG257</i>	TTTCGAGTTCCTTCGCCAGGTCTCATCC	AAGGATGAGACCTGGCGAAGGAACCTCGA
<i>AsE246</i>	TTCAGGGAGTGCTTCAGCATTAGCGTTG	GTGTTGAATCCATCCATGACTGAGGCAG
<i>AsIB259</i>	GGAAGGAATGGCAACAGCAGCAGAAAGA	TGTTCTATGGTTGATGTTTTGCTTCAAATGCA

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, *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

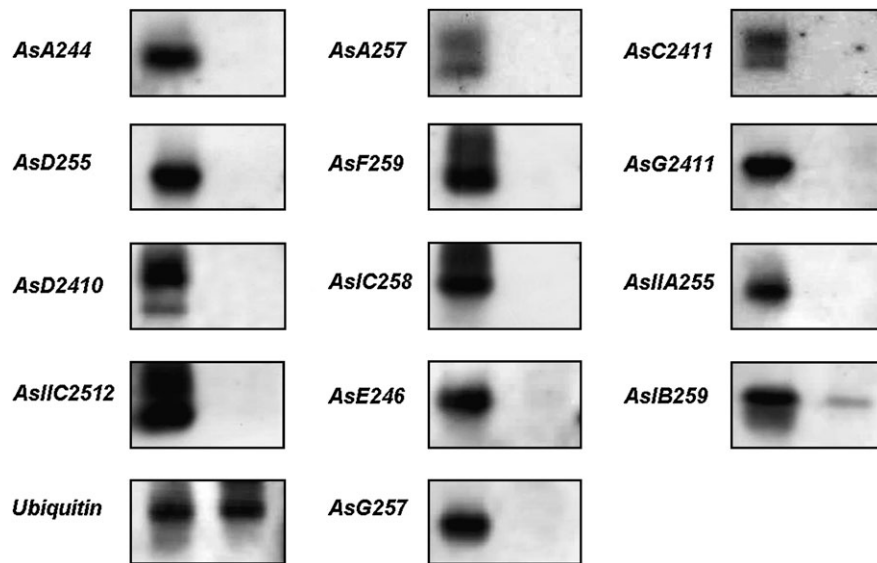


Fig. 1. Virtual northern blot analysis for the genes predominantly or exclusively expressed in the infected roots. The left column represents the infected roots and the right column represents the uninfected roots. Each sample of SMART cDNAs was resolved in a 1.2% (w/v) agarose gel, transferred to a Hybond N⁺ nylon membrane, and hybridized with radiolabelled probes. Ubiquitin is the loading control.

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 AsA257, AsC2411, AsF259, AsG2411, AsIC258, AsIIA255, and AsIIC2512, contained two pairs of cysteines, whereas the second group involving AsA244, AsG257, 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

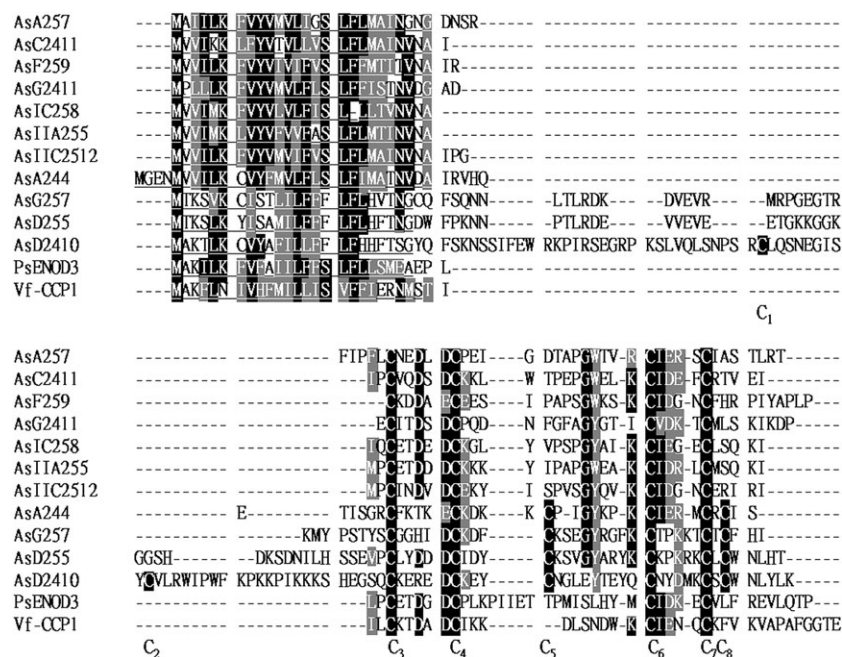


Fig. 2. Alignment of the deduced polypeptide sequences of *Astragalus sinicus* late nodulins AsA257, AsC2411, AsF259, AsG2411, AsIC258, AsIIA255, AsIIC2512, AsA244, AsG257, AsD255, and AsD2410, the early pea nodulin PsENOD3 (AAB23537), and *Vicia faba* late nodulin Vf-CCP1 (CAB96472). Identical residues 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.

Table 3. Sequence analysis of nodule-specific or nodule-enhanced genes revealed by SSH from *A. sinicus*

Gene	Accession no.	Length of putative polypeptide (no. of amino acids)	Best homology (BLASTP, NCBI) and E value	Domain analysis with Pfam (trusted matches) and E value
AsA244	DQ199634	70	Hypothetical protein, <i>Galega orientalis</i> CAB51773, 0.002	Nodulin late, 1e-08
AsA257	DQ199641	68	ENOD3, <i>Pisum sativum</i> AAB23537, 0.18	Nodulin late, 1.5e-08
AsC2411	DQ199638	60	ENOD3, <i>Pisum sativum</i> AAB23537, 0.23	Nodulin late, 9e-14
AsF259	DQ199639	64	Late nodulin, <i>Vicia faba</i> CAB96475, 0.036	Nodulin late, 7.2e-11
AsG2411	DQ199636	63	Hypothetical protein, <i>Galega orientalis</i> CAB51773, 2e-04	Nodulin late, 1.4e-13
AsIC258	DQ199640	58	Late nodulin, <i>Vicia faba</i> CAB96475, 0.007	Nodulin late, 6.7e-14
AsIIA255	DQ199635	59	ENOD3, <i>Pisum sativum</i> AAB23537, 0.002	Nodulin late, 7.3e-11
AsIIC2512	DQ199637	62	Late nodulin, <i>Vicia faba</i> CAB96475, 0.006	Nodulin late, 4.2e-14
AsD255	DQ199650	101	No hits below 1e	
AsD2410	DQ199646	126	No hits below 1e	
AsG257	DQ199643	90	No hits below 1e	Defesin 2, 1.2e-05
AsE246	DQ199648	132	NLTP, <i>Lycopersicon esculentum</i> CAJ19705, 2e-15	Tryp_alpha_amyl, 1.9e-11
AsIB259	DQ199649	148	Unknown protein <i>Arabidopsis thaliana</i> AAL66904, 6e-11	Tryp_alpha_amyl, 0.0071
AsB2510	DQ199647	148	Leghaemoglobin, <i>Sesbania rostrata</i> CAA31859, 3e-54	

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

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.

```

GGGCCATAGACGTGGAAAATAATGACTAAAGTCTCAAGTATATTTGCAATGATTCCTCTTTTCTATT 75
G P * T W K I M T K S L K Y I S A M L L F F F L F

CTTCATTTCAACCAATGGTGGTTTCCAAAAACAATCCGACTTTGAGGGATGAAGTGGTGAAGTAGAGGAA 150
L H F T N G D W F P K N N P T L R D E V V E V E E

ACGGGAAAAAAGGGGAAAGGGGAGCCATGATAAATCAGATAACATTCGTAAAGCTTATCATTCAAGCCCTC 225
T G K K G G K G G S H D K S D N I L

TTTTCATTCTCTTTTATAATACATTATTTTATTTTAAAAATACCTTTATTCGTGTTTAAATATATTGATA 300
TTTATCCATTTTTACATTACATGTCAGTGCATGACGAGTTCATGCTTTATGACGATGACTGTATAGAC 374
H S S E V P C L Y D D D C I D

TATTGCAAAAGTGTGGCTATGACGCTTATAAGTGTAAACCGAAAGGAAGTGTCTGCTGGAACCTCATACT 449
Y C K S V G Y A R Y K C K P K R K C L C W N L H T

TAATGCTATTAACTGAGTAATAACTTACTTATCGTACAAATATGATTGTCTCTTATTATTTTCTCTAATCATT 524
* *

TTGTAGTCTGCTTTCAACAGACATTCATTAATCGGGTCAACGGGTTTGACCTGTCAAGTCCAGAGGGAAGT 599
GCTAGTTGACTCTAGGAATCATTTCAGTTTGGGATTCGACAGTCTAAAACTTATGATGTTTGGAGCAAGC 674
AAATGCTTCGTTTCGACACAGACTGAACATGTGAACGTGGGCGAGATAGTTATTTTCAAACTTCTGCGGT 749
CACTTAGTTTTAAATTTTCTCTCATTTGTAACTAGTATAAATAAGAGTTTTCCTGATGGAACCTTGTGCATT 824
TTCACACTAGTTTCTTAAGCTTTAATCTCCACACATATCGTGAGACACGAACACAGACTGCGAATGTA 899
TGAACCTGCGTACCACCTTGAAGCAATTTACTGTTTCTCTAAAAATCTGCAAAAAAAAAAAAAAAAAAAAA 974
AAAAAAGTA 983

```

Fig. 3. Genomic DNA and translated sequence of the *AsD255* clone. The putative amino acid sequence is shown with a predicted signal peptide (dotted line) and six cysteine residues (bold characters). One intron interrupting the *AsD255* coding region and another beyond the coding region are underlined. The start codon and two proximate stop codons are in bold.

Table 4. Comparison of the putative polypeptides of 13 genes identified

Primary polypeptide	Signal peptide		Processed polypeptide		
	No. of amino acids	pI	No. of amino acids	pI	kDa
AsA244	30	4.37	40	9.42	4.72
AsA257			68	4.66	7.53
AsC2411	26	9.70	34	4.44	3.40
AsF259			64	6.52	7.24
AsG2411	26	5.59	37	4.04	4.00
AsIC258	25	8.34	33	4.36	3.63
AsIIA255	26	8.34	33	6.02	3.82
AsIIC2512	26	8.34	36	4.86	4.03
AsD255	24	9.70	77	6.98	8.82
AsD2410	24	9.19	102	9.13	12.27
AsG257	24	9.31	66	8.94	7.60
AsE246	23	5.75	109	8.66	12.10
AsIB259	28	8.50	120	9.15	13.54

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 an intron 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* (Kajjalainen *et al.*, 2002). Kato *et al.* (2002) characterized five CCP genes with reduced expression in ineffective

A

AsE246(As)	-----	-MKFAYVVVV	M-----	-CI	MVVLNPSMTE	AETISCREVV	VTLTPCFPYL	LSGYG-PSQS	CCEAIKSFKI
LTP(Le)	-----	---MEMVNKI	A-----	-CFV	LLCMVVVAPH	AEALTCQVVT	STLAPCLPYL	MNRG--PLRN	CCDGVKGLLG
LPT1(EI)	MLKLTHIQS	HITRNI IYNN	AGIKVAVLVV	ALMVVASQMY	ANAITCQQVS	SSLAPCVNFL	KSGGA-PSQP	CQNGLGGMVN	
LTP2(Lp)	-----	---MEMVNKI	A-----	-CFV	LLCMVVVAPH	AEALTCQVVT	STLAPCLPYL	MNRG--PLGG	CCGQVKGKLLG
NLTP(Nt)	-----	---MEMVSKI	A-----	-CFV	VLCMVVAPH	AEALTCQVQ	SSLAPCVPYL	LGRG--PLGG	CCGQVKGKLLG
LTP(An)	-----	-MASSVVKL	A-----	-CAV	FMCMVLAAPH	AEALTCQVVT	SSMTPOMSYL	TGGGS-PTPA	CCGQVKSLSNS
LT1(Os)	-----	-MA	RAQ-LVLVAV	V-----	-AAL	LL---AAPH	AVAITCQVNV	SAVGPCLTYA	RG-GAGPSAA
LTP(Pc)	-----	-MASSAVIKL	A-----	-LVV	ALCMAVS--V	AHAITCQSVS	ANLAPCINIV	RSGGA-VPPA	CQNGIKTING
LTP(Mx)	-----	-MASSAVINL	A-----	-LVV	ALCMAVS--V	AHAITCQSVS	SNLVPDFDYV	RSGGP-VPPA	CQNGIRTING
LTP(Lc)	-----	---MEMVNKI	A-----	-CFV	LLCMVVVAPH	AEALTCQVVT	STLAPCLPYL	MNRG--PLGG	CCGQVKGKLLG
LTP4(Ng)	-----	-MAKVALLV	V-----	-CMV	AVAVMLTPHA	DAATSCQVNV	TSLTPCINIV	RQGGG-IPAA	CCSGIKTSLNS
PLTP(Zm)	-----	-MA	RTQSAVAVAV	V-----	-AAV	LLAAAAATTS	EAAITCQVVS	SAIAPCLSYA	RGTGSGPSAS
NLTP(Ec)	-----	-----	-----	-----	-----	-AISCQVVS	SAIGPCLAYA	RGAGAAPSAS	CQSGVRSLSNA
Consensus						C V	PC		CC

AsE246(As)	VFKNKINQI	ACNC-MKAA	--FFGLSNAN	AEALPEKCNV	KMHYKINTSF	DCTSIOQLKN	VNVEKIQILQ	TLLV	
LTP(Le)	QAKTTVDRQA	ACTC-LKSAA	SSFTGLNLGK	AAALPNTCSV	NIPYKISPSST	DCKVQ----			
LPT1(EI)	QAKSTADKQA	ACNC-LKTAA	KNMPGLNPAN	AESLPSCCKV	NIPYKISFST	NONSIK----			
LTP2(Lp)	QAQTTVDRQT	ACTC-LKSAA	SSFTGLDLGK	AASLPSTCSV	NIPYKISPSST	DCKVQ----			
NLTP(Nt)	AARTPADRKCT	ACNC-LKSAA	NFTKGIDMGN	AARLPCTCGV	NIPYKISPSST	DCKVQ----			
LTP(An)	MASTPADRKA	ACGC-LKSAA	GAMTNLMGN	AASLPCKGCI	SLPYPISTST	DCKVN----			
LT1(Os)	AASSTADRRRT	ACNC-LKNAA	RGIKGLNAGN	AASIPSKCGV	SVPYTISASI	DCSRVS----			
LTP(Pc)	LAKTTPDRQA	ACNC-LKNLA	GSVSGVNPON	AESLPCKCGV	NWPKISTST	NCATVK----			
LTP(Mx)	LAKTTPDRQA	ACNC-LKSLA	GSVSGVNPON	AESLPCKCGV	NWPKISTST	NCATVK----			
LTP(Lc)	QAQTTVDRQA	ACAC-LKSAA	SSFTGLDLGK	AASLPSTCNV	NIPYKISPSST	DCKVQ----			
LTP4(Ng)	QAQSTPDRQT	ACNC-LKSAA	RSISGINFGL	AASLPCKCGV	NIPYKISPSI	DCKTVQ----			
PLTP(Zm)	AASTAADRRRA	ACNC-LKNAA	RGVSGLNAGN	AASIPSKCGV	SIPYTIISTST	DCSRVN----			
NLTP(Ec)	AARTTADRRRA	ACNCSLKSA	SRVSGLNAGK	ASSIPCKCGV	RLPYAISASI	DCSRVN----			
Consensus		AC C K A		A P C	Y I S C				

B

AsIB259(As)	MVMNFVTSF	IITMVMILMA	SRVLVLEADL	IWPHCGR-TY	KSVARSINPC	LDAAHNMTN	VPTICCSMD	VLLQ-MQPLC	
Unknown1(At)	---MKCKKF	VAVALMSLLI	SLASVEAAG-	---ECGRMPI	NQAAASLSPC	LPATKNPRGK	VPPVCCAKVG	ALIR-TNPRC	
OJ991214(Os)	-MRSSPIIFL	LLAAAAIAA	PWQVAHAAG-	---KCGKTPA	EKVALKLAPC	AKAAQDPGAR	PAAOCCAAVR	DIGHQSHAC	
LTP(At)	-METHVLSRI	FLVL-CIYS	LKTMVYGAG-	---ECGRNPP	DREAIKLAPC	AMAAQDTSK	VSAICCARVK	QMG--QNPKC	
Unknown(Os)	-MRAGRLVIL	VAVAVVVVVA	GGAAGAEAG-	---ECGRASA	DRVALRLAPC	VSAADDPOSA	PSSSCCSAVH	TIG--QSPSC	
Unnamed(At)	-METHVLSRI	FLVL-CIYS	LKTMVYGAG-	---ECGRNPP	DREAIKLAPC	AMAAQDTSK	VSAICCARVK	QMG--QNPKC	
Unknown2(At)	-----MK	FLCILGLILL	IGTVVDGAG-	---ECGRSSP	DNEAMKLAPC	AGAAQDANSA	VPGGCTQIK	RFS--QNPKC	
LTP(At)	---MAASMK	FLCILGLILL	IGTVVDGAG-	---ECGRSSP	DNEAMKLAPC	AGAAQDANSA	VPGGCTQIK	RFS--QNPKC	
Consensus				A	CG	A L PC	A	CC	C

AsIB259(As)	LCETLLSYIA	IKPEIILSAA	VAIPSRONLH	YPYNIKKCAI	VVP-SPKRGK	IPNPGREFPR	HLTRPETEWR	K	
Unknown1(At)	LCAVMLSPLA	KKAGINPGIA	IGVPKRKNIR	NRPAKRCRG-	-----RYI	VP-----			
OJ991214(Os)	LCAVLLSSTV	RRSGVKPEVA	ITIPKRCCLA	NRPVGYKCG-	-----AYT	LPSLQG----			
LTP(At)	LCAVMLSSTA	RSSGAKPEIS	MTIPKRONIA	NRPVGYKCG-	-----AYT	LP-----			
Unknown(Os)	LCAVMLSNTA	RVAGIKPEVA	ITIPKRONMA	DRPVGYKCG-	-----DYT	LP-----			
Unnamed(At)	LCAVMLSSTA	RSSGAKPEIS	MTIPKRONIA	NRPVGYKCG	NDIKSLTFS	ISLQQLFCLS	NVFNLFSCSL	YSALKQOTES	
Unknown2(At)	LCAILLSDTA	KASGVDPEVA	LTIPKRONFA	NRPVGYKCG-	-----AYT	LP-----			
LTP(At)	LCAILLSDTA	KASGVDPEVA	LTIPKRONFA	NRPVGYKCG-	-----AYT	LP-----			
Consensus	LC	LS		P RC	C				

C

AsE246(As)	-----	-MKFAYVVVV	MCIMVVLNPS	MTEAETISCR	EVVVTLTPCF	PYLLSGYGPS	---QSCCEAI	KSFKI	VFKNK
AsIB259(As)	MVMNFVTSF	IITMVMILMA	SRVLVLEADL	IWPHCGRTYK	SVARSINPCL	DAAHNMTNV	P--TICCSMV	DVLLQ----	
NLTP(Mb)	-----	-----	-----	-----	MTQG	QVQGNLAQC1	GFLOKGGVVP	---PSCCTGV	KNILNSSRIT
LTP(Ps)	-----	---MSKVLLIS	VMTACMLISS	SYGKATLTCE	QVTIWLTPC1	PYGTLGGSVL	---PLCCQGV	HSLNAAKNG	
NLTP(Ps)	-----	-----	---MVVSAP	MAEA-AISCG	AVTAAPVPCF	AYLKGAPSPS	---LQCCQGV	RRLNG----	
LTP(Ms)	-----	-----	-----	---SCG	TVTGALSPC1	AYLRGGAGPS	---PACCAGV	KRLNAAATTT	
MtN5(Mt)	-----	-MA	HSQKALAQW	MIGALLFAML	AGSLAVQ1CN	IDPNDLKSC-	SKFVTGRNPP	RADEACCGVL	R-----
LTP1(Vr)	-----	-MASIKCACV	VALLCLVAVT	APTAHAITCG	QVASSLTSC1	PFITKGGIVP	---PSCCAGV	KSLNAAAKTT	
AKCS9(Vu)	-----	-MTMKM	KMKMSVVCV	VVALFLIDV	GPVAEAVTCN	--PTELSSCV	PAITGGSKPS	---STCCSKL	K-----
LTP(Ca)	-----	-MASMKVVCV	ALIMCIVIAV	MAES-AITCG	RVDITALAPCL	GYLOGGPGPS	---AQCCQGV	RNLNSAAVTT	
Consensus					C	C		CC	

AsE246(As)	INGQIACNM	KKAA--FFGL	SNANAELPE	KCNVKNHYKI	NTSFDCTS1Q	DLKNVVEKI	QILQTLV--	-----	
AsIB259(As)	--MQPLCLCE	TLISYIAIKP	EIILSAAVAI	PSRNLHYPY	NIKK-CAIVV	PSPKRGKIPN	PGRFEPRLT	RPETEWK	
NLTP(Mb)	ADRRVACSCL	KAAAGAVRG1	NPNNAEALPG	KCGVNIPIYKI	STSTNCNCS1	-----	-----	-----	
LTP(Ps)	DDRLACHCV	QDRAALPLI	DYTRINQIGD	LQSKCPHKV	YPSIDCDKVK	-----	-----	-----	
NLTP(Ps)	PDRKGVNCL	KGAAGSVPLG	KPGNVAAALPA	NVLDLFLSRL	APLPTMPSV	FKKKCG----	-----	-----	
LTP(Ms)	PDRQAACNCL	KSAAGATISGL	NASTAAGLPG	KCGVNIPIYKI	ST-----	-----	-----	-----	
MtN5(Mt)	-RANLPCLCG	YKSALTYYG1	NAKKALALPG	QCGLQTP----	---SNC	-----	-----	-----	
LTP1(Vr)	PDRQAVNCL	KSEAGRIIGF	NANNAAILPG	KCGVSIPIYKI	STSTNCASIK	F-----	-----	-----	
AKCS9(Vu)	--VOEPCLCN	YIKNPSLKQY	VNSPGAKKVL	SNCGVITYP--	---NC	-----	-----	-----	
LTP(Ca)	PDRQAACNCL	KSAAGTSIRL	NANNAALALPG	KCGVNIPIYKI	STSTNCATIR	V-----	-----	-----	
Consensus		C C		C	C				

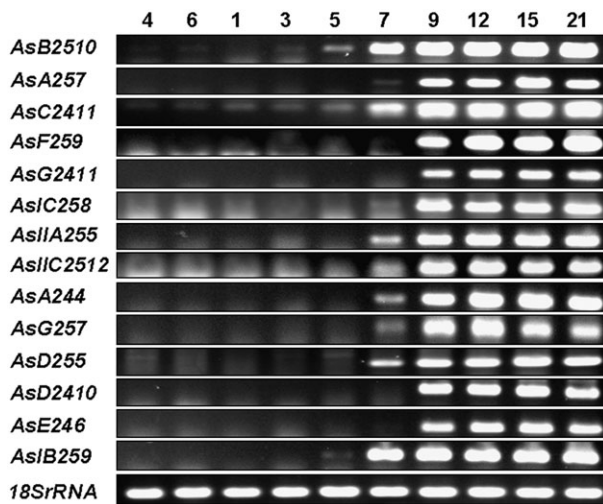


Fig. 5. RT-PCR analysis of mRNA for the 13 genes under investigation in the total RNA isolated from: 4, 4-d-old uninfected roots; 6, 6-d-old uninfected roots (harvested just before inoculation); 1, infected roots 1 d post-inoculation (dai); 3, infected roots at 3 dai; 5, infected roots at 5 dai; 7, infected roots at 7 dai; 9, infected roots at 9 dai; 12, infected roots at 12 dai; 15, infected roots at 15 dai; 21, infected roots at 21 dai. As a symbiotic marker characteristic of functional nodules, amplification of *A. sinicus* *Lb* AsB2510 was included. The 18S rRNA was amplified by RT-PCR as a constitutive control.

nodules on *P. sativum* and thought that their successive expression during nodule development was associated with nitrogen-fixing activity. Crockard *et al.* (2002) isolated 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 *et al.* (2003) and by Graham *et al.* (2004) using computational approaches. Further, Graham *et al.* pointed out that CCPs were not specific to galeoid legumes and functioned in indeterminate 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

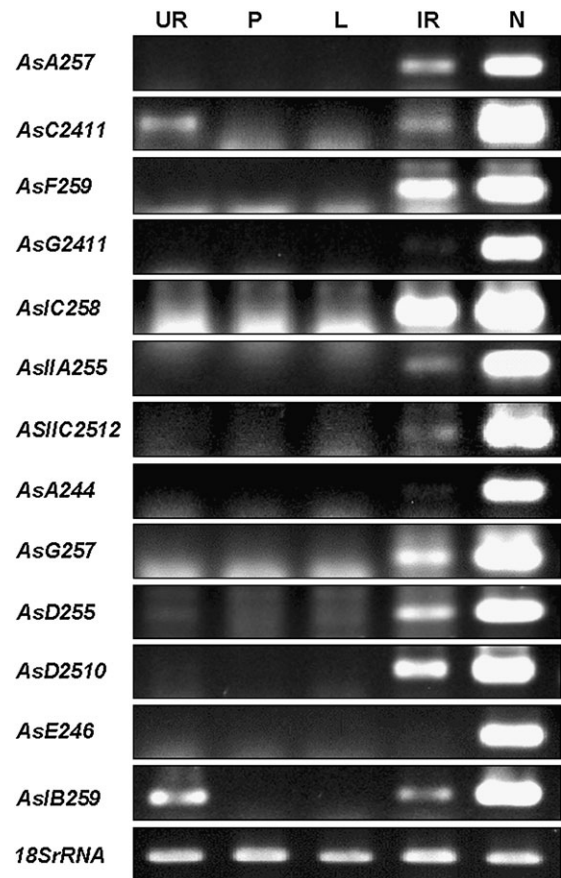


Fig. 6. RT-PCR analysis of mRNA for the 13 genes under investigation in the total RNA isolated from: UR, 33-d-old uninfected roots; P, petioles at 27 dai; L, leaves at 27 dai; IR, infected roots with nodules removed at 27 dai; N, nodules at 27 dai. The 18S rRNA was amplified by RT-PCR as a constitutive control.

et al., 2003; Graham *et al.*, 2004). As the molecular effectors of innate immunity, defensins are widely distributed 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 α/β motif stabilized by eight cysteine residues and composed of three antiparallel β -sheets and one α -helix (Almeida *et al.*, 2002). This kind of structural pattern was also found in insect defensins and some scorpion

Fig. 4. Alignment of the deduced amino acid sequences of AsE246 and AsIB259 with LTPs or LTP-like proteins from various plants. Compared with AsE246 and AsIB259, identical residues conserved in at least 50% of the sequences are on a grey background. The signal peptides of AsE246 and AsIB259 are underlined. The consensus sequence is based on the identical residues conserved in all sequences, except cysteine. (A) Comparison between the deduced sequence of AsE246 and the LTPs from various plants (CAJ19705, AAM00272, AAB07487, AAM74206, BAC77694, AAP97429, AAF26451, AAT8066, AAZ22829, AAT68263, AAB06443, and P23802). (B) Comparison between the deduced sequence of AsIB259 and the unknown proteins or LTPs from *Arabidopsis thaliana* and *Oryza sativa* (AAK96826, CAE01523, NP_568160, XP_475631, BAB10801, AAM65889, and NP_850700). (C) Comparison of the deduced sequences of AsE246 and AsIB259 with the LTPs from legumes (ISIIY_A, AAF61436, CAA74892, AAZ32875, MTY15371, AAQ74627, Q43681, and CAA05771). As, *Astragalus sinicus*; Le, *Lycopersicon esculentum*; El, *Euphorbia lagascae*; Lp, *Lycopersicon pennellii*; Nt, *Nicotiana tabacum*; An, *Atriplex nummularia*; Os, *Oryza sativa*; Pc, *Pyrus communis*; Mx, *Malus domestica*; Lc, *Lycopersicon chilense*; Ng, *Nicotiana glauca*; Zm, *Zea mays*; Ec, *Eleusine coracana* (finger millet); At, *Arabidopsis thaliana*; Mb, mung bean (*Vigna radiata* var. *radiata*); Ps, *Pisum sativum*; Ms, *Medicago sativa*; Mt, *Medicago truncatula*; Vr, *Vigna radiata*; Vu, *Vigna unguiculata* (cowpea); Ca, *Cicer arietinum*.

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 intra-chain 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-X₄-Cys-X₇-Cys-X₄-Cys-X₁-Cys motif of AsA244. For example, the arthropod defensin family including insect and scorpion cysteine-rich antibacterial peptides have a Cys-X_n-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. truncatula* (Gamas *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 elicitor 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 AsIIC2512, 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 micro-organisms 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

- Almeida MS, Cabral KM, Kurtenbach E, Almeida FC, Valente AP. 2002. Solution structure of *Pisum sativum* defensin 1 by high resolution NMR: plant defensins, identical backbone with different mechanisms of action. *Journal of Molecular Biology* **315**, 749–757.
- Bateman A, Coin L, Durbin R, *et al.* 2004. The Pfam protein families database. *Nucleic Acids Research Database* issue **32**, D138–D141.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783–795.
- Blein JP, Coutos-Thévenot P, Marion D, Ponchet M. 2002. From elicitors to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. *Trends in Plant Science* **7**, 293–296.
- Boman HG. 2003. Antibacterial peptides: basic facts and emerging concepts. *Journal of Internal Medicine* **254**, 197–215.
- Brewin NJ. 1991. Development of the legume root nodule. *Annual Review of Cell Biology* **7**, 191–226.
- Bulet P, Cociancich S, Reuland M, Sauber F, Bischoff R, Hegy G, Van Dorsselaer A, Hetru C, Hoffmann JA. 1992. A novel insect defensin mediates the inducible antibacterial activity in larvae of the dragonfly *Aeschna cyanea* (Plecoptera, Odonata). *European Journal of Biochemistry* **209**, 977–984.
- Cammue BPA, Thevissen K, Hendriks M, *et al.* 1995. A potent antimicrobial protein from onion seeds showing sequence homology to plant lipid transfer proteins. *Plant Physiology* **109**, 445–455.
- Carvalho AO, Teodoro CES, Cunha MD, Okorokova-Facanha AL, Okorokov LA, Fernandes KVS, Gomes VM. 2004. Intracellular localization of a lipid transfer protein in *Vigna unguiculata* seeds. *Physiologia Plantarum* **122**, 328–336.
- Chen HK, Shu MK. 1944. Notes on the root nodule bacteria of *Astragalus sinicus* L. *Soil Science* **58**, 291–293.
- Chen KC, Lin CY, Kuan CC, Sung HY, Chen CS. 2002. A novel defensin encoded by a mungbean cDNA exhibits insecticidal activity against bruchid. *Journal of Agricultural and Food Chemistry* **50**, 7258–7263.
- Chen WX, Li GS, Qi YL, Wang ET, Wang HL, Yuan HL, Li JL. 1991. *Rhizobium huakuii* sp. Nov. isolated from the root nodules of *Astragalus sinicus*. *International Journal of Systematic Bacteriology* **41**, 275–280.
- Choi DW, Song JY, Oh MH, Lee JS, Moon J, Suh SW, Kim SG. 1996. Isolation of a root-specific cDNA encoding a ns-LTP-like protein from the roots of bean (*Phaseolus vulgaris* L.) seedlings. *Plant Molecular Biology* **30**, 1059–1066.
- Colilla FJ, Rocher A, Mendez E. 1990. γ -Purothionins: amino acid sequence of two polypeptides of a new family of thionins from wheat endosperm. *FEBS Letters* **270**, 191–194.
- Cornet B, Bonmatin JL, Hetru C, Hoffmann JA, Ptak M, Vovelle F. 1995. Refined three-dimensional solution structure of insect defensin A. *Structure* **3**, 435–448.
- Crockard A, Bjourson J, Dazzo B, Cooper JE. 2002. A white clover nodulin gene, dd23b, encoding a cysteine cluster protein, is expressed in roots during the very early stages of interaction with *Rhizobium leguminosarum* biovar *trifolii* and after treatment with chitolipooligosaccharide Nod factors. *Journal of Plant Research* **115**, 439–447.
- De Samblanx GW, Goderis IJ, Thevissen K, Raemaekers R, Fant F, Borremans F, Acland DP, Osborn RW, Patel S, Broekaert WF. 1997. Mutational analysis of a plant defensin from radish (*Raphanus sativus* L.) reveals two adjacent sites

- important for antifungal activity. *Journal of Biological Chemistry* **272**, 1171–1179.
- El Yahyaoui F, Kuster H, Ben Amor B, et al. 2004. Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. *Plant Physiology* **136**, 3159–3176.
- Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP. 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiology* **130**, 519–537.
- Frühling M, Albus U, Hohnjec N, Geise G, Pühler A, Perlick AM. 2000. A small gene family of broad bean codes for late nodulins containing conserved cysteine clusters. *Plant Science* **152**, 67–77.
- Fujie M, Nakanishi Y, Murooka Y, Yamada T. 1998. *AsNODc22*, a novel nodulin gene of *Astragalus sinicus*, encodes a protein that localizes along the cell wall of bacteria induced cells in a nodule. *Plant and Cell Physiology* **39**, 846–852.
- Gamas P, Niebel Fde C, Lescure N, Cullimore J. 1996. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. *Molecular Plant-Microbe Interactions* **9**, 233–242.
- García-Olmedo F, Molina A, Segura A, Moreno M. 1995. The defensive role of non-specific lipid-transfer proteins in plants. *Trends in Microbiology* **3**, 72–74.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The proteomics protocols handbook*. Totowa, NJ: Humana Press Inc., 571–608.
- Graham MA, Silverstein KAT, Cannon SB, VandenBosch KA. 2004. Computational identification and characterization of novel genes from legumes. *Plant Physiology* **135**, 1179–1197.
- Györgyey J, Vaubert D, Jimenez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E. 2000. Analysis of *Medicago truncatula* nodule expressed sequence tags. *Molecular Plant-Microbe Interactions* **13**, 62–71.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Research, Symposium Series* **41**, 95–98.
- Hanks JN, Snyder AK, Graham MA, Shah RK, Blaylock LA, Harrison MJ, Shah DM. 2005. Defensin gene family in *Medicago truncatula*: structure, expression and induction by signal molecules. *Plant Molecular Biology* **58**, 385–399.
- Hirsch AM. 1992. Developmental biology of legume nodulation. *New Phytologist* **122**, 211–237.
- José-Estanyol M, Gomis-Rüth FX, Puigdomènech P. 2004. The eight-cysteine motif, a versatile structure in plant proteins. *Plant Physiology and Biochemistry* **42**, 355–365.
- Kader JC. 1996. Lipid-transfer proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**, 627–654.
- Kajjalainen S, Schroda M, Lindstrom K. 2002. Cloning of nodule-specific cDNAs of *Galega orientalis*. *Physiologia Plantarum* **114**, 588–593.
- Kato T, Kawashima K, Miwa M, Mimura Y, Tamaoki M, Kouchi H, Suganuma N. 2002. Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. *Molecular Plant-Microbe Interactions* **15**, 129–137.
- Kobayashi Y, Takashima H, Tamaoki H, et al. 1991. The cysteine stabilized α -helix: a common structural motif of ion channel blocking neurotoxic peptides. *Biopolymers* **31**, 1213–1220.
- Krause A, Sigrist CJ, Dehning I, Sommer H, Broughton WJ. 1994. Accumulation of transcripts encoding a lipid transfer-like protein during deformation of nodulation-competent *Vigna unguiculata* root hairs. *Molecular Plant-Microbe Interactions* **7**, 411–418.
- Lindsey K, Casson S, Chilley P. 2002. Peptides: new signalling molecules in plants. *Trends in Plant Science* **7**, 78–83.
- Liu KH, Lin TY. 2003. Cloning and characterization of two novel lipid transfer protein I genes in *Vigna radiata*. *DNA Sequence* **14**, 420–426.
- Madsen EB, Madsen LH, Radutoiu S, et al. 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**, 637–640.
- Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK. 2002. A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. *Nature* **419**, 399–403.
- Malek W, Inaba M, Ono H, Kaneko Y, Murooka Y. 1998. Competition for *Astragalus sinicus* root nodule infection between its native microsymbiont *Rhizobium huakuii* bv. Renge B3 and *Rhizobium* sp. ACMP18 strain, specific for *Astragalus cicer*. *Applied Microbiology and Biotechnology* **50**, 261–265.
- Melo FR, Ridgen DJ, Franco OL, Mello LV, Ary MB, Grossi-de-Sa MF, Bloch CJ. 2002. Inhibition of trypsin by cowpea thionin: characterization, molecular modeling, and docking. *Proteins: Structure, Function, and Genetics* **48**, 311–319.
- Mendez E, Moreno A, Colilla F, Pelaez F, Limas GG, Mendez R, Soriano F, Salinas M, de Haro C. 1990. Primary structure and inhibition of protein synthesis in eukaryotic cell-free system of a novel thionin, gamma-hordothionin, from barley endosperm. *European Journal of Biochemistry* **194**, 533–539.
- Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E. 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiology* **132**, 161–173.
- Mithofer A. 2002. Suppression of plant defence in rhizobia-legume symbiosis. *Trends in Plant Science* **7**, 440–444.
- Mitra RM, Long SR. 2004. Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula*/*Sinorhizobium meliloti* symbiosis. *Plant Physiology* **134**, 595–604.
- Naito Y, Fujie M, Usami S, Murooka Y, Yamada T. 2000. The involvement of a cysteine proteinase in the nodule development in Chinese milk vetch infected with *Mesorhizobium huakuii* subsp. *rengei*. *Plant Physiology* **124**, 1087–1095.
- Nap JP, Bisseling T. 1990. Developmental biology of a plant-prokaryote symbiosis: the legume root nodule. *Science* **250**, 948–954.
- Nielsen H, Krogh A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. In: Glasgow J, Littlejohn T, Major F, Lathrop R, Sankoff D, Sensen C, eds. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, 122–130.
- Park HC, Kang YH, Chun HJ, et al. 2002. Characterization of a stamen-specific cDNA encoding a novel plant defensin in Chinese cabbage. *Plant Molecular Biology* **50**, 57–68.
- Pastorello EA, Pompei C, Pravettoni V, Brenna O, Farioli L, Trambaioli C, Conti A. 2001. Lipid transfer proteins and 2S-albumins as allergens. *Allergy* **56**, 45–47.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Research* **33**, W116–W120.
- Sanchez-Monge R, Gomez L, Barber D, Lopez-Otin C, Armentia A, Salcedo G. 1992. Wheat and barley allergens associated with baker's asthma. Glycosylated subunits of the

- alpha-amylase-inhibitor family have enhanced IgE-binding capacity. *Biochemical Journal* **281**, 401–405.
- Scheres B, van Engelen F, van der Knaap E, van de Wiel C, van Kammen A, Bisseling T.** 1990. Sequential induction of nodulin gene expression in the developing pea nodule. *The Plant Cell* **2**, 687–700.
- Silverstein KA, Graham MA, Paape TD, VandenBosch KA.** 2005. Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiology* **138**, 600–610.
- Sterk P, Booiij H, Schellekens GA, Van Kammen A, De Vries SC.** 1991. Cell-specific expression of the carrot EP2 lipid transfer protein gene. *The Plant Cell* **3**, 907–921.
- Terras FR, Eggermont K, Kovaleva V, et al.** 1995. Small cysteine-rich antifungal proteins from radish: their role in host defense. *The Plant Cell* **7**, 573–588.
- Thoma S, Kaneko Y, Somerville C.** 1993. A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. *The Plant Journal* **3**, 427–436.
- Thomma BPHJ, Cammue BPA, Thevissen K.** 2002. Plant defensins. *Planta* **216**, 193–202.
- VandenBosch KA, Stacey GA.** 2003. Summaries of legume genomics projects from around the globe. Community resources for crops and models. *Plant Physiology* **131**, 840–865.
- Vasse J, de Billy F, Camut S, Truchet G.** 1990. Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *Journal of Bacteriology* **172**, 4295–4306.
- Verdoy D, Lucas MM, Manrique E, Covarrubias AA, de Felipe MR, Pueyo JJ.** 2004. Differential organ-specific response to salt stress and water deficit in nodulated bean (*Phaseolus vulgaris*). *Plant, Cell and Environment* **27**, 757–767.
- Yamada K, Natori S.** 1993. Purification, sequence and antibacterial activity of two novel sapecin homologues from *Sarcophaga* embryonic cells: similarity of sapecin B to charybdotoxin. *Biochemical Journal* **291**, 275–279.