

## Identification of a New Inducible Nodulation Gene in *Azorhizobium caulinodans*

Koen Goethals, Peter Mergaert, Mengsheng Gao, Danny Geelen, Marc Van Montagu, and Marcelle Holsters

Laboratorium voor Genetica, Universiteit Gent, B-9000 Gent, Belgium.

Received 20 May 1992. Accepted 20 July 1992.

The narrow host range bacterial strain *Azorhizobium caulinodans* ORS571 induces the formation of nitrogen-fixing nodules on the root and stem of the tropical legume *Sesbania rostrata*. Here, a new flavonoid-inducible locus of ORS571 is described, locus 4. The locus was identified and isolated via the occurrence of particular sequences, the  $\gamma$  and  $\delta$  elements. These elements are reiterated in the ORS571 genome, linked to symbiotic loci. Sequencing of locus 4 showed the presence of an open reading frame (ORF6) that is flanked downstream by a  $\gamma$  element and upstream by a  $\delta$  element. The  $\gamma$  element is approximately 180 bp in size, and shows homology to the insertion element *ISRM3*, an insertion sequence belonging to a distinct class of IS elements. The  $\delta$  element is about 300 bp in size and has homology with

repeated sequences found in other Rhizobiaceae. The ORF6 gene product shows a low, but significant homology to the mouse mastocytoma antigen P35B (Szikora *et al.*, EMBO J. 9: 1041-1050, 1990) and to a class of NAD/NADP-binding sugar epimerase/dehydrogenases (Pissowtzki *et al.*, Mol. Gen. Genet. 231: 113-213, 1991). Immediately upstream from ORF6, a *nod* box-related sequence is present, the arrangement of which is fully consistent with a recently presented model for the *nod* box structure (Goethals *et al.*, Proc. Natl. Acad. Sci. USA 89: 1646-1650, 1992). Insertional inactivation of ORF6 did not affect the nodulation and fixation performance on *S. rostrata*. However, on *S. formosa* roots the nodulation kinetics of such a mutant was clearly affected (about 5 days delay). We propose to call this new symbiotic gene *nolk*.

The genera *Rhizobium* and *Bradyrhizobium*, together with the more recently described genus *Azorhizobium* (type species *A. caulinodans* strain ORS571; Dreyfus *et al.* 1988) constitute a group of soil bacteria that are capable of inducing the formation of nitrogen-fixing nodules on the roots of leguminous plants. The host plant for *A. caulinodans* is the tropical legume *Sesbania rostrata* Brem which grows during the rain season in the Sahel region of West Africa and, as an adaptation to waterlogged conditions, carries vertical rows of dormant root primordia all along its stem. These primordia will develop into roots when immersed in water; they develop into nitrogen-fixing nodules upon infection by *A. caulinodans* (Dreyfus and Dommergues 1981). A characteristic of strain ORS571 is its narrow host range of infection: Only on *S. rostrata* fully nitrogen-fixing nodules are formed. On other *Sesbania* species nodules may be induced that are, however, ineffective ( $\text{Fix}^-$ ) (Ndoye 1990).

In *Bradyrhizobium* and *Rhizobium* species sets of nodulation (*nod*) genes have been identified that are involved in the nodulation process and are organized in *nod* operons under the positive transcriptional control of one or more regulatory *nodD* genes (for a review, see Long 1992). These *nodD* genes encode LysR-type regulatory proteins (NodD) that respond to signal molecules present in host plant exudates. Upon expression of the *nod* operons, the bacteria produce a return signal (the *nod* factor), a lipo-oligosac-

charide of specific structure that is released in the environment and induces changes related to the nodulation process in the host plant cells (Lerouge *et al.* 1990; Spaik *et al.* 1991; Truchet *et al.* 1991; Dénarié and Roche 1992). In *A. caulinodans*, an essential *nod* locus (locus 1) carries genes homologous to the (brady)rhizobial common *nodABC* genes and is transcribed in the presence of *S. rostrata* seedling exudate (Van den Eede *et al.* 1987; Goethals *et al.* 1989). The most prominent inducer molecule in this exudate was shown to be 7,4'-dihydroxyflavanone (liquiritigenin; Messens *et al.* 1991). Also, the related flavanone naringenin turned out to be a good inducer of the azorhizobial *nodABC* operon (Goethals *et al.* 1989). The induction process is mediated by the azorhizobial locus 3, encoding the NodD activator protein (Goethals *et al.* 1990) that binds to conserved motifs (NodD boxes) found in *nod* promoter sequences (*nod* boxes) (Goethals *et al.* 1992).

As an approach to studying *nod* factor biosynthesis, we want to identify and sequence the bacterial genes that are expressed early in the interaction, under the control of the *nodD* gene. The identification of these genes in *A. caulinodans* is hampered by the absence of a Sym plasmid and by the fact that the symbiotic loci found until now are dispersed over the chromosome. Moreover, sequencing of several ORS571 *nod* genes revealed them to be too divergent from (brady)rhizobial homologues to be picked up by DNA hybridizations.

Another way to proceed is to make use of repeated sequences that often seem to be linked to symbiotic loci (reviewed by Martinez *et al.* 1990). Here, we give an example of such an approach: The isolation of a new plant-inducible *nod* gene is described, which was identified by the presence of flanking repeated sequences that are also located upstream from locus 1 and downstream from locus 3.

Present address of M. Gao: Institute for Application of Atomic Energy, P.O. Box 510g, Beijing, People's Republic of China.

Nucleotide and/or amino acid sequence data are to be submitted to GenBank, EMBL, and DDBJ as accession number JO3713.

## MATERIALS AND METHODS

**Strains and plasmids.** The strains and plasmids used are listed in Table 1. Growth conditions and antibiotic concentrations for ORS571 and *Escherichia coli* were as described before (Goethals *et al.* 1989).

**Molecular biology techniques.** Standard molecular biology techniques for DNA isolation, restriction analysis, DNA labeling, cloning, Southern blotting, and hybridization were as described by Sambrook *et al.* (1989). Low-stringency hybridizations were carried out at 60° C using 6× SSC (1× SSC is 0.18 M NaCl plus 0.015 M sodium citrate). Electroporation of *E. coli* cells was done with a gene pulser (Bio-Rad, Richmond, CA) according to the manufacturer's instructions.

Insertion mutagenesis of the locus 4 was achieved by cloning of the  $\Omega$  cassette of pPH45 $\Omega$  (see Table 1) in the unique *Bss*HII site of pRG9011 (carrying itself the 2.3-kb *Sal*I subfragment of pRG901 cloned in pBR325; Table 1) yielding pRG9011- $\Omega$ B. Simultaneously, the 8.4-kb *Eco*RI insert fragment of pRG901 was subcloned in the *Eco*RI site of the broad host range vector pLAFRI (Table 1) yielding pRG9010, and this construct was electroporated into the *polA*<sup>-</sup> *E. coli* strain CSH2110. Next, pRG9011- $\Omega$ B was electroporated to CSH2110(pRG9010) and, as pBR325 cannot replicate in a *polA*<sup>-</sup> background, growth on a medium containing tetracycline (pLAFRI), spectinomycin, and streptomycin ( $\Omega$ ) resulted in the selection of single recombinants in which pRG9010 and pRG9011- $\Omega$ B are cointegrated. Double recombinants, whereby the  $\Omega$  cassette is introduced in the 8.4-kb *Eco*RI fragment, were isolated after screening for the loss of the pBR325 carbenicillin marker and were inspected by restriction enzyme analysis (data not shown). From a resulting recombinant, pRG9010- $\Omega$ B, the 10.4-kb *Eco*RI insert fragment was isolated and recloned in the *Eco*RI site of pBR325, generating construct pRG901 $\Omega$ B. Finally, this plasmid was mobilized by triparental mating to ORS571 and, because pBR325 cannot replicate in this strain, screening for tetracycline-sensitive and streptomycin/spectinomycin-

resistant colonies resulted in the isolation of homogenotes where the wild-type DNA is exchanged for the mutant DNA as controlled by hybridization analysis (data not shown). The mutant strain ORS571- $\Omega$ S, carrying an  $\Omega$  element in the *Sma*I site upstream from *nolK* (Fig. 1) was generated in a similar way.

DNA sequencing procedures were basically as described by Sanger *et al.* (1977). Sequence processing and interpretation was done with the IG Suite and PC/GENE packages from IntelliGenetics, Inc., Mountain View, CA. Oligonucleotide synthesis was carried out on an Applied Biosystems (Foster City, CA) model 394 DNA/RNA synthesizer according to the manufacturer's recommendations.

**Triparental matings and  $\beta$ -galactosidase tests.** Triparental mating using pRK2013 as a helper plasmid (Table 1) and quantitative  $\beta$ -galactosidase assays using *o*-nitrophenyl  $\beta$ -D-galactoside as a substrate were carried out as described before (Goethals *et al.* 1989).

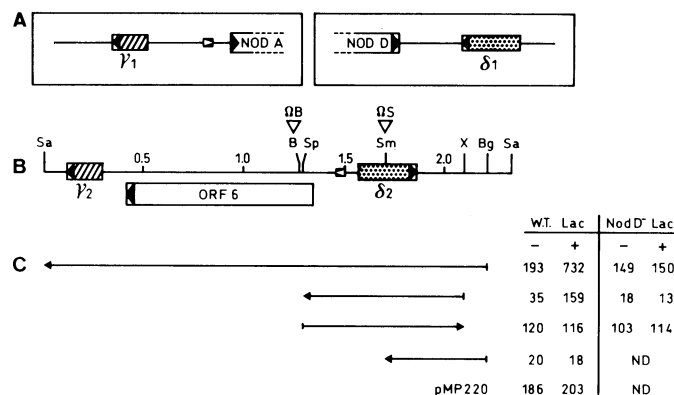
**Nodulation assays.** Sterilization of *S. rostrata* seeds, growth of *S. rostrata* seedlings, nodulation, and acetylene reduction assays were done according to Van den Eede *et al.* (1987). For the other plants tested, essentially the same approach was followed. Nodulation was scored every 2 days. For every strain tested, 10 individual plants were inoculated; each experiment was done at least twice.

## RESULTS

**Identification of *nod* locus 4.** The expression of *A. caulinodans nod* locus 1 is naringenin inducible, and the

**Table 1.** Strains and plasmids used in this study

	Relevant characteristics	Reference
<b>Strains</b>		
MC1061	<i>E. coli</i> strain, <i>araD139</i> $\Delta$ ( <i>ara leu</i> ) $\Delta$ <i>lacX74 galU galK hsr hsm<sup>+</sup> strA</i>	Casadaban and Cohen (1980)
CSH2110	<i>E. coli</i> strain, <i>polA</i> <sup>-</sup> , <i>Nal</i> <sup>R</sup>	Heffron <i>et al.</i> (1977)
ORS571	<i>Azorhizobium caulinodans</i> type strain able to nodulate stems and roots of <i>Sesbania rostrata</i>	Dreyfus <i>et al.</i> (1988)
ORS571-3	A <i>nodD</i> <sup>-</sup> derivative of ORS571	Goethals <i>et al.</i> (1989)
<b>Plasmids</b>		
pBR325	repColE1 Cb <sup>R</sup> Tc <sup>R</sup> Cm <sup>R</sup>	Bolivar (1978)
pRK2013	repColE1 Km <sup>R</sup> Tra <sup>+</sup> <i>mob</i> <sup>+</sup> used as helper plasmid in conjugations	Ditta <i>et al.</i> (1980)
pLAFRI	Wide host range cosmid, Tc <sup>R</sup> Tra <sup>-</sup> <i>mob</i> <sup>+</sup>	Friedman <i>et al.</i> (1982)
pPH45 $\Omega$	Plasmid carrying the Sm/Sp $\Omega$ cassette	Prentki and Krisch (1984)
pMP220	Promoter probe plasmid containing a promoterless LacZ gene, Tc <sup>R</sup>	Spaink <i>et al.</i> (1987)



**Fig. 1.** Physical-genetic map of locus 4. **A**, Schematic representation of the locus 1 upstream region (left) and the locus 3 downstream region (right). Striped box,  $\gamma_1$  element; dotted box,  $\delta_1$  element. Black arrowheads indicate the direction of translation of the respective genes or the arbitrary orientation of a repeated element. An open arrowhead indicates the position and the polarity of a *nod* box sequence. Scale is the same as for **B**. **B**, Physical-genetic map of the locus 4. Black arrowheads point the direction of ORF6 (white box) expression or the relative orientation of the repeated elements  $\gamma$  and  $\delta$ . An open arrowhead indicates the position and orientation of a *nod* box sequence.  $\Omega$ B and  $\Omega$ S indicate the positions of the  $\Omega$  cassette insertions in the respective mutants (see Results). Striped box,  $\gamma_2$  element; dotted box,  $\delta_2$  element. Abbreviations; B, *Bss*HII; Bg, *Bg*III; Sa, *Sal*I; Sm, *Sma*I; Sp, *Sph*I; X, *Xho*I. Scale is in kilobases. **C**, Representation of the different fragments (relative to **B**) cloned in pMP220. The arrows indicate the orientation of the *lacZ* gene. The table shows the  $\beta$ -galactosidase units (Miller units, see Materials and Methods) for the respective cloned fragment in a wild type ORS571 (W. T.) and in the *nodD*<sup>-</sup> strain ORS571-3 in the presence (+) and absence (-) of 10  $\mu$ M naringenin. ND, not determined.

locus contains a *nod* box-type promoter (NBI) (Goethals *et al.* 1992). To identify other similarly regulated ORS571 genes an oligonucleotide spanning the NBI was synthesized, end-labeled, and used as a probe in hybridizations under different stringencies against several digests of total ORS571 DNA. Apart from the expected bands, no other related sequences could be identified using this approach. However, when a 12.8-kb *EcoRI* fragment containing part of locus 1 and approximately 8 kb upstream (Goethals *et al.* 1989) was used as a probe, many fragments from an *EcoRI*-digested ORS571 total DNA cross-hybridized, even at high stringencies. This observation was taken as an indication of the presence of repeated sequences in the ORS571 genome. A preliminary characterization of cross-hybridizing sequences showed the occurrence of several classes of repeats, some of which are present in only two copies, others in five or more (M. Holsters, unpublished results). One particular element, called  $\gamma$ , was present in two copies only, one copy located immediately upstream of the *nodABC* operon ( $\gamma_1$ ) (Fig. 1), another copy located on an 8.4-kb *EcoRI* fragment ( $\gamma_2$ ). This 8.4-kb *EcoRI* could

be isolated as part of a cosmid clone upon colony hybridization against an ORS571 pLAFRI genomic library. It turned out that the same 8.4-kb *EcoRI* fragment cross-hybridized to another sequence  $\delta_1$ , present downstream from the ORS571 *nodD* gene in locus 3. Like the  $\gamma$  elements, the  $\delta$  elements are present in only two copies in the ORS571 genome. On the 8.4-kb fragment,  $\gamma_2$  and  $\delta_2$  are separated by approximately 1.5 kb. The linkage of the  $\delta_1$  and  $\gamma_1$  elements to ORS571-symbiotic loci prompted us to investigate the possible occurrence of symbiotic genes, perhaps similarly regulated as locus 1, linked to  $\delta_2$  and  $\gamma_2$ . A screening for naringenin-inducible promoter activity was done by cloning of subfragments of the 8.4-kb *EcoRI* fragment in the promoter probe plasmid pMP220 (Table 1), upstream from a promoterless  $\beta$ -galactosidase gene and determining the levels of *lacZ* expression in *Azorhizobium* in the presence and the absence of naringenin. As shown in Figure 1, an inducible promoter activity could indeed be ascribed to an approximately 2.3-kb *SalI*-*BglII* fragment and was further delimited to be contained within an approximately 400-bp *SphI*-*SmaI* fragment. This inducible

```

GTCGACGAGGATGATCTGCAGCGCCGAGGCTGGCTGATCGTGCATGACACGGCTGCCGAAGAAGGGAGAGCGTTTCGGTGGCGTGCCTCCGCAATACGCTCGGCCCTTAGCAAGATCTCAACTGCCAGACCCCTCGTATC
GCTGACGCTGGCGGCCGACGGCCGACGGCCGACGGCCGACGGCTGTACCCTGTTCCGCGTGTCCACAGTTGACAGGAGGCAACTTGCTCGAGCCATCGACGGCTACGAACAGGTAGAGCTCGCCCTCGGGGCTGTAACCTCGG
CGATGTCGATGTGGAAGACACCGATCGGGTAGGATTTGAACCGCTTCTTCGCTCCTTGTGCGCTGCGATCTCTGGTAGCCGGCTAATGCCCTGGGGCTGCAGGCAGCAATTCAGGGACGTGCCCGTCAGATCGGGATCGTGC
CCTGACGCGGTACAGCAATCGTCCCGCGGACGCAACGTGTGGTGGCCGAAAGGCGACGATGATCGCTCTTCCTCGACAGACAGAACCGTTGACTCGGGGCTTCGGGGCGGTGGCAGGTCGGCAACTGAATTGCGCTGCT
TCCACCTGGGGGCTGTTTTCTGGTTAAATCCCGTCGCGCTTCGCGGGCCCTTAGGCTCTTGTACTGTGCTGATCGCTCGACGGACCGCATCTGCTGCTGTTGGCGACCGATGGAGAACCCTGGCCATAGCGCATCTTCCA
TTCTGCGAAAAGACTGACAGTCAAAATCCGGATCAAACTAGACTCGCGTCCGAAAGTACGAAACCGAAGTACCCGTTGGGTCAGGCTAAAATCGAAAATTCAAAATCGGGCTATCCATCGGATGGATAATTCGGATAATCA
AAAAGGATTTTGTTTTTTAAATTCACACAGTATCGCGAGAAGTGGGCATTAACAAGGTCAGAC
ATG CGC GAG AGC GAA GGG CTT TAC GGC GAC AGG GAG GGA ATC CCC CGC [GTG] GGT AAA GGT AAA AAA CTC TTG ATC ACC GGT GGT CGA GGT ATG GTC GGG CGA
MET Arg Glu Ser Glu Gly Leu Tyr Gly Asp Arg Glu Gly Ile Pro Arg [Val] Gly Lys Gly Lys Lys Leu Leu Ile Thr Gly Gly Arg Gly MET Val Gly Arg
AAC TTG ATT GCA TGC GCG GCG CGC TCG GGC TGG GAA ATA ATA GCG CCC ACC TCA GTC GAT CTG GAT TTG CGG AAT GCA GAG GCG GTC GAG CAA TAT ATC CGT
Asn Leu Ile Ala Cys Ala Ala Arg Ser Gly Trp Glu Ile Ile Ala Pro Thr Ser Val Asp Leu Asp Leu Arg Asn Ala Glu Ala Val Glu Gln Tyr Ile Arg
CGT CAG CTA CCG GAT GTT GTG GTT CAC GCC GCT GGC GTA GTC GGT GGC ATC CAC GCA AAT ATC GCC GAT CCG ATA CAT TTT CTC GCA GAT AAT GCG GCG ATG
Arg Gln Leu Pro Asp Val Val Val His Ala Ala Gly Val Val Thr Leu Ile Asn Leu Ser Ser Ser Cys MET Tyr Pro Ala Cys Ile Glu Gly Pro Leu Lys
GCT CTA AAT GTC GTC ATG TCG AGC TTT CGT TCG GAA GTG GTT ACA TTG ATC AAC TTG TCC TCG AGT TGC ATG TAT CCT GCC TGC ATC GAG GGG CCT TTA AAA
Ala Leu Asn Val Val MET Ser Ser Phe Arg Ser Glu Val Val Thr Leu Ile Asn Leu Ser Ser Ser Cys MET Tyr Pro Ala Cys Ile Glu Gly Pro Leu Lys
GAG TGC GAC ATC CTG CGT GGC CCG TTT GAA GTG ACT AAT GAA GGA TAT GCT CTA GCG AAG ACG GTT GGT CTT AAG ATT TGT GAA TAC ATT GAC AAG CTG CCG
Glu Cys Asp Ile Leu Arg Gly Pro Phe Glu Val Thr Asn Glu Gly Tyr Ala Leu Ala Lys Thr Val Gly Leu Lys Ile Cys Glu Tyr Ile Asp Lys Leu Pro
AAT TTT AAT TAC AAA ACA CTG ATT GCA TGC AAT TTA TAT GGA GTG GGC GAC AAT TTT GAT CCA AGA AGA TCG CAT CTT CTG CCT GCA ATA ATA GAA AAA ATT
Asn Phe Asn Tyr Lys Thr Leu Ile Ala Cys Asn Leu Tyr Gly Val Gly Asp Asn Phe Asp Pro Arg Arg Ser His Leu Leu Pro Ala Ile Ile Glu Lys Ile
CAC AAA GCC TCT CAA TGT GGT AGC GAA AGT GTG AGT ATC TGG GGC GAT GGG ACT GCT AGG CGT GAG TTT ATG TTC GCG TAC GAT TTT GCA AAA ATT ATC ATT
His Lys Ala Ser Gln Cys Gly Ser Glu Ser Val Ser Ile Trp Gly Asp Gly Thr Ala Arg Arg Glu Phe MET Phe Ala Tyr Asp Phe Ala Lys Ile Ile Ile
AAG GCA TTG GAA GTT CCG GAA TTA ATA CCT AGT TCA ATG AAT GTC GGT GTT GGG AAA GAT TTG TCT GTG TTA GAA TAT TAC AGC TTG GTG GCT CGT GTA ATT
Lys Ala Leu Glu Val Pro Glu Leu Ile Pro Ser Ser MET Asn Val Gly Val Gly Lys Asp Leu Ser Val Leu Glu Tyr Tyr Ser Leu Val Ala Arg Val Ile
GGG TGG AGC GGT GAA TTT GTC TAT GAC TTG AAT AGG CCT GTG GGT ATG CGT AGT AAA TTG ATG GAT ATC ACG CAI CTT ACG GCA TTA GGT TGG GTC CCA GAA
Gly Trp Ser Gly Glu Phe Val Tyr Asp Leu Asn Arg Pro Val Gly MET Arg Ser Lys Leu MET Asp Ile Thr His Leu Thr Ala Leu Gly Trp Val Pro Glu
CGG TCG CTT GAA GGG GGG ATT CGA TCG ACA TAC CAG TAT TAT ATC ACC GGG AAT GAG GTC TAC GAG TAG AGTCGTTCTATCGGTGCTCTGCTGGGGAATGGGGTCTCGGTCGGG
Arg Ser Leu Glu Gly Gly Ile Arg Ser Thr Tyr Gln Tyr Tyr Ile Thr Gly Asn Glu Val Tyr Glu ***
GAGACCTGAGTGGCCCTTCAGGATTTGACATTCGGGTCGATGACGTAGACGGCACCGAATTCATAGGGGAGTGTGAGGAATTCGTGTGCGGGGCGCCATAGTGGACCTGATAGAGGTTTCCGATGACGACGCAAGGCCAC
CTCGTAGTCTGACGCGCTTCTGGATCAGCTTCTGGCCGGCCGATCCCAAGACGGCCCTTTGACACCAATGGGCTGCTCGATGATCTGAAGAAGCGCTGGCCAAGCGGGCGCTGAAGGGCAATTTATCGCGCCGTGGACGCA
CGGCCGGCGAGATCGCGCTCGGGGCCCTTCGAGGAAAGCCTGTTGGGCTCCGCTACCCGGCTATAGCCAGAGCTGTCGAC

```

Fig. 2. Sequence of the locus 4 2,333-bp *SalI* fragment. The repeated sequences  $\gamma_2$  and  $\delta_2$  are underlined. The orientation of the fragment is opposite to that of Figure 1. The boxed sequence upstream of ORF6 is the *nod* box-related sequence. Immediately upstream of the ATG and GTG possible start codons, the potential Shine-Dalgarno sequences are underlined. The potential GTG start codon is boxed. Three asterisks indicate the ORF6 stop codon.

expression was dependent on the presence of an intact *nodD* gene (Fig. 1). As it turned out from mapping of homologies and sequencing data (see further), the inducible promoter activity should not be ascribed to either of the two repeated elements but to a *nod* box-type sequence.

**Sequence of *nod* locus 4.** The sequence of the entire 2.3-kb *SalI* fragment was determined and is presented in Figure 2. An open reading frame (ORF6 in Fig. 1) starts either at an ATG start codon or at a GTG codon located 48 bp further downstream. In the former case (ORF6<sub>ATG</sub>), a 329-amino acid protein would be encoded with a molecular mass of 36,225 Da and a pI of 5.88. In the latter case (ORF6<sub>GTG</sub>), a 313-amino acid protein of molecular mass 34,378 Da and a pI of 6.08 would be produced. We favor the latter start codon (GTG) as the actual *in vivo* start codon because it is preceded by a more elaborate Shine-Dalgarno (SD) sequence (Stormo *et al.* 1982). Furthermore, using the PC/GENE computer program for detection of protein-coding regions, a higher coding score was obtained for ORF6<sub>GTG</sub> (score 8) than for ORF6<sub>ATG</sub> (score 7). Databank (SWISSPROT release 20 and PIR release 30) searches with the ORF6 gene product as a query revealed a low, but significant homology with the mouse mastocytoma transplantation antigen P35B (Szikora *et al.* 1990) (Fig. 3). Further examination of the sequence revealed the presence of a putative NAD/NADP-binding domain in the amino-terminal part of the protein. Low, but significant, homology was found with a class of nucleotide sugar metabolizing enzymes that bind NAD/NADP and of which UDP-glucose-4-epimerase (Lemaire and Müller-Hill 1986) showed the best fit (Fig. 3; for more details, see Discussion).

Upstream from ORF6, in a 422-bp *SmaI-SphI* fragment associated with the inducible promoter activity, a correctly oriented *nod* box-related sequence is present. Upon comparison of this *nod* box (NB2) with *nod* boxes from (brady)rhizobia and with the ORS571 *nod* box from *nod* locus 1 (NB1), a high sequence divergence was found in all cases (about 50% homology; Fig. 4). The sequence



**Fig. 3.** Homology of the ORF6 gene product with the mouse mastocytoma antigen P35B and UDP-glucose-4-epimerase. a, P35B antigen; b, ORF6 gene product; c, UDP-glucose-4-epimerase (*Escherichia coli*). Double points indicate identical amino acids or conserved substitutions belonging to either of the groups EQDN, WYF, HKR, or AGST. Amino acid positions identical in the three sequences are shaded. The putative  $\beta$ -turn- $\alpha$ -helix- $\beta$ -turn motif found in NAD/NADP-interacting proteins is boxed. The invariant glycine residues are indicated separately in bold.

conservation between NB1 and NB2 is fully consistent with a model proposed by us concerning the fundamental structure of a *nod* box (see Discussion).

The start codon of the *nodA* gene in *nod* locus 1 is located 92 bp downstream of the NB1 3' end. The GTG start codon of the ORF6<sub>GTG</sub> is located at a comparable 98 bp downstream from the NB2 3' end. In the case of the ATG start of ORF6<sub>ATG</sub>, this distance is only 50 bp. This again argues for the GTG of ORF6<sub>GTG</sub> being the actual translational start site for ORF6.

Sequence comparison of the ORF6-flanking regions with the *nod* locus 1-upstream region (Goethals *et al.* 1989) and the *nod* locus 3-downstream region (D. Geelen, unpublished results) revealed the nature and extent of the  $\gamma$  and  $\delta$  repeated sequences, respectively. The alignment of the two  $\gamma$  and the two  $\delta$  sequences is shown in Figure 5. The  $\gamma$  element encompasses about 180 bp, with 84% homology; the  $\delta$  elements are approximately 300 bp in size and are 95% homologous. Neither element contains direct or inverted repeats.

**Mutagenesis of *nod* locus 4.** To investigate the role of the plant-inducible ORF6 in the interaction between the bacteria and the plant, we constructed a mutant derivative by inserting the  $\Omega$  cassette in a cloned fragment in *E. coli* and homogenotizing the mutation in the ORS571 genome (see Materials and Methods). The  $\Omega$  fragment was isolated from pPH45 $\Omega$  (Prentki and Krisch 1984; Table 1) and carries a gene for streptomycin and spectinomycin resistance flanked at both sides by transcriptional and translational stop signals. The cassette was inserted in the *Bss*HIII site located at the 5' end of ORF6, and the construct was homogenotized in the ORS571 genome. The resulting homogenote (ORS571- $\Omega$ B) was tested in comparison to the wild-type ORS571 for root and stem nodulation behavior on *S. rostrata*. No significant differences were observed—neither in nodulation kinetics nor in nodule numbers. Acetylene reduction levels in the nodules of both strains were the same (data not shown). Some additional root inoculation experiments were carried out to check the competitiveness of the mutant strain. Coinoculations of *S. rostrata* roots with a 1:1 mixture of ORS571 wild type and ORS571- $\Omega$ B showed that the same numbers of the

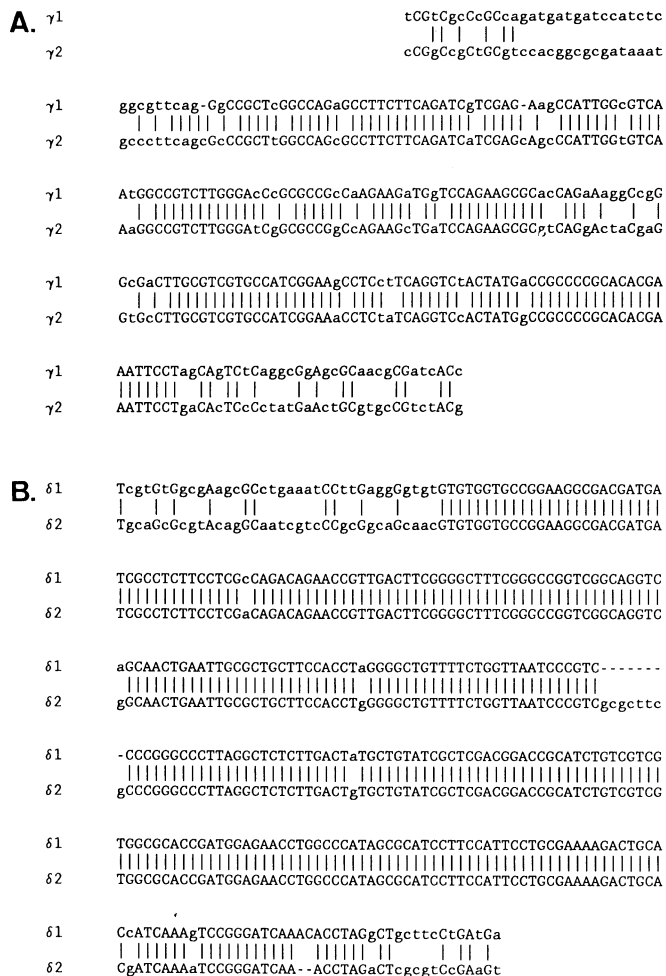


**Fig. 4.** The *nod* box-related sequence of locus 4. **A**, Alignment of the *nod* boxes of: Rm(abc), the *R. meliloti* common *nod* operon; Ac(ORF6), the *A. caulinodans* locus 4 (NB2); Ac(abc), the *A. caulinodans* common *nod* operon (NB1). The boxed bases indicate the NodD boxes: the repeat ATC-N<sub>9</sub>-GAT or a related motif. **B**, Indication of the inverted repeat structure of the NodD boxes found in the *nod* box of: Rm(abc), the *R. meliloti* common *nod* operon; Ac(ORF6), the *A. caulinodans* locus 4 (NB2); Ac(abc), the *A. caulinodans* common *nod* operon (NB1). A point indicates the center of the palindrome.

two types of bacteria were present in the nodules. We also used a dilution series of wild type and mutant to nodulate roots and observed no differences.

To test the possible involvement of ORF6 in host specificity of nodulation, we compared the nodulation behavior of wild type and mutant on several tropical legumes, some of which allow the formation of non-nitrogen-fixing ( $\text{Fix}^-$ ) nodules with ORS571. The plants tested were *S. grandiflora*, *S. cannabina*, *S. emerus*, *S. formosa*, *Acacia albida*, *A. radiana*, *Aeschynomene indica*, and *Neptunia oleracea*. The only difference in nodulation behavior occurred on root nodulation of *S. formosa*. On this *Sesbania* species, *A. caulinodans* can induce large, white, nonfixing nodules. An approximate 5-day delay in nodule appearance was observed with ORS571- $\Omega$ B over a 25-day period (Fig. 6). Because of this phenotype on the one hand and the expression pattern on the other hand, we propose to call the described locus a *nod* locus (*nod* locus 4) and propose the name *nolK* for ORF6.

Using a *nolK*-internal *Bss*HIII-*Stu*I fragment as a probe for hybridization against ORS571 total DNA, we observed some faintly hybridizing bands at lower stringencies (data



**Fig 5.** Sequence alignment of the  $\gamma$  and  $\delta$  elements. **A**, Sequence alignment of the  $\gamma_1$  and  $\gamma_2$  repeated elements. Some flanking bases are included to show the divergence at the borders. **B**, Sequence alignment of the  $\delta_1$  and  $\delta_2$  repeated elements. Some flanking bases are included to show the divergence at the borders.

not shown). These data might indicate the presence of different *nolK*-related genes in ORS571, a finding that could explain the weak effect of the *nolK* mutation on nodulation.

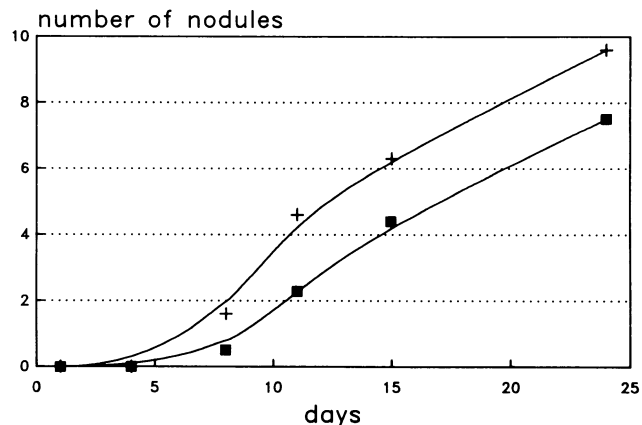
An ORS571 derivative with an  $\Omega$  insertion in the  $\delta_2$  element of locus 4 (ORS571- $\Omega$ S; Fig. 1) was also constructed and tested on *S. rostrata* and the other plant species listed above. No differences in nodulation or nitrogen fixation characteristics were observed.

## DISCUSSION

In this paper, a new symbiotic locus of *A. caulinodans* strain ORS571 is described. The expression of this locus (locus 4) is induced by the flavanone naringenin. This induction is mediated by the *A. caulinodans nodD* gene product (NodD) and could be correlated with the presence of a *nod* box-related sequence (NB2). The sequence of NB2 displays only about 50% homology to (brady)rhizobial *nod* boxes as well as to the *A. caulinodans nod* box from the common *nod* locus (NB1 in locus 1). This might indicate that locus 1 and locus 4 were acquired independently from different strains, or it might illustrate the ancient nature of the ORS571 *nod* genes.

The nucleotide conservation between NB1 and NB2 is in concordance with a recently proposed model concerning the fundamental structure of a *nod* box (Goethals *et al.* 1992). In this model, the *nod* box as originally defined (Rostas *et al.* 1986) represents a modular structure with different juxtaposed NodD boxes acting as the binding sites for the NodD proteins and having the typical inverted repeat structure ATC-N<sub>9</sub>-GAT. Also, in NB2 two such consecutive NodD boxes occur (Fig. 4). The most 5'-located NodD box of NB2 constitutes a perfect inverted repeat over its entire length (Fig. 4).

Downstream from NB2, an open reading frame (ORF6) is present, the derived product of which showed a low, but significant homology with the mouse mastocytoma transplantation antigen P35B (Szikora *et al.* 1990), the function of which is unknown. In the amino-terminal part of the ORF6 protein, a motif could be discerned that corresponds to the so-called  $\beta$ -turn- $\alpha$ -helix- $\beta$ -turn domain



**Fig. 6.** Nodulation kinetics on *Sesbania formosa*. +, ORS571 nodulation; ■, ORS571- $\Omega$ B nodulation. For each strain, the data are the mean of 10 plants tested.

involved in NAD/NADP binding by several proteins (Rossman *et al.* 1975). Typical for this domain is the presence of a stretch of invariant glycine residues. As these motifs are characteristically located at the very amino-terminus of a protein, this again argues for the GTG start codon being the actual translation start. A low, but significant, homology was also found between the ORF6 gene product and a class of sugar epimerases that use NAD/NADP as cofactor and of which UDP-glucose-4-epimerase (GalE) of *E. coli* (Lemaire and Müller-Hill 1986) was taken as an example (Fig. 3). The amino acid conservation between the ORF6 gene product and both the P35B and the UDP-glucose-4-epimerase protein, corresponds well with the amino acid residues noted by Pissowatzki *et al.* (1991) to be more or less conserved between nucleotide sugar metabolizing enzymes. Therefore, we postulate that the ORF6 gene product encodes a nucleotide sugar epimerase/dehydrogenase that uses NAD/NADP as a cofactor.

Insertional inactivation of ORF6 resulted in a derivative strain (ORS571- $\Omega$ B) that was able to nodulate the stems and the roots of *S. rostrata* equally well as the wild type ORS571. Also, no differences in fixation capacities of the nodules were evident. Different other *Sesbania* species (which form Fix<sup>-</sup> nodules with ORS571) and other genera of tropical Leguminosae were used in root nodulation tests to investigate the possible involvement of the locus in host range determination. The only difference scored was a reduced nodulation kinetics on *S. formosa*; nodules appeared on this plant with an approximate 5-day delay. This phenotype suggests the involvement of ORF6 in nodule induction and, combined with the expression pattern, led us to assign the locus as a *nod* locus and to propose the name *nolK* for ORF6. Possibly, the locus is involved in the optimization of the nodulation process under conditions occurring in the natural environment, the rain season in the Sahel region, and as such functions in the ecology of the nodulation process.

Hybridization studies suggested the occurrence of other *nolK*-related sequences in ORS571. Perhaps this could explain the weak effect of *nolK* inactivation on the nodulation behavior as shown before for the *R. meliloti nodPQ* and *nodM* genes (Schwedock and Long 1990; Baev *et al.* 1991).

The *nolK* gene is flanked at both sides by sequences that are repeated once in the ORS571 genome, linked to other symbiotic loci. The  $\gamma$  sequence is approximately 180 bp long with one copy present downstream of *nolK* and the other copy (84% homologous) located upstream of the common *nod* locus of ORS571. Databank searches with this region as a query revealed an intriguing homology with *ISRm3*, an insertion sequence that was described in *R. meliloti* and that is a member of a distinct class of IS elements (Wheatcroft and Laberge 1991). The *ISRm3* homology extends 280 bp to the left of  $\gamma_1$  (Fig. 1) until the end of the available sequence information. The degree of homology adds up to an overall 50%, but in the last approximate 200 bp of the sequence this increases up to 68%. At the protein level, a clearcut homology (about 60%) with the *ISRm3* transposase is evident when one ignores three frameshift mutations in the 5' end of the  $\gamma_1$ -linked sequence. These data imply the occurrence in *A. caulinodans*

*dans* of two related IS elements (or their relics), belonging to the *ISRm3* family. It is tempting to speculate on the possible involvement of these elements in the evolution and dispersion of the ORS571 nodulation genes.

Upstream of *nolK*, the 300-bp  $\delta_2$  element is present, a 95% homologous copy of which ( $\delta_1$ ) is located downstream of the ORS571 *nodD* gene. When the  $\delta$  element was used as a query against a databank, we found some 53% homology with a *R. fredii* repetitive element that is associated with the *nodABC* genes of this strain (Krishnan and Pueppke 1991). The homology was confined to a part of this repetitive element carrying two overlapping ORFs, which are also found in the *Agrobacterium rhizogenes* Ri T-DNA, although at the protein level no obvious homology could be shown. Whereas the *R. fredii* repetitive element is homologous to the *B. japonicum* RSR $\alpha$  insertion sequences, we could only detect a weak homology between the  $\delta$  element and RSR $\alpha$  sequences. Again the true nature of the  $\delta$  elements remains obscure, although one can argue that they are derived from insertion elements and may have played (or still play) a role in the evolution of the nodulation genetics of *A. caulinodans*.

Currently, we are investigating the structure and biosynthesis of the azorhizobial *nod* factor. In the future, we would like to correlate structural properties of the factor with the function of specific *nod* genes. As a first approach we can study the effect of the *nolK* mutation on structural modifications of the ORS571 *nod* factor.

Isolation of other ORS571 *nod* genes involved in factor biosynthesis is impeded by the absence of a Sym plasmid, the divergence of the *nod* sequences compared to the ones of (brady)rhizobia and their dispersed localization on the ORS571 chromosome. Furthermore, the divergence of the *nod* box sequence in *A. caulinodans* itself precludes the possibility of using a *nod* box probe as a tag for flavonoid-inducible *nod* genes. In fact, hybridization experiments with such a probe gave no useful results. As an alternative, we can make use of the peculiarity of some repeated elements to be linked to symbiotic loci. As an illustration, we showed the isolation of *nod* locus 4 (*nolK*) using the  $\gamma$  element. Still other classes linked to symbiotic loci are present in the ORS571 genome and may be exploited further to clone other genes involved in nodulation and eventually factor biosynthesis.

#### ACKNOWLEDGMENTS

We thank B. Dreyfus for providing seeds of the tested legumes and help with the nodulation assays, B. Baum for indicating the *ISRm3* homology with the  $\gamma$  region, and the Department of Plant Molecular Biology (Leiden University, The Netherlands) for providing plasmid pMP220. We further thank Martine De Cock for layout and Vera Vermaercke and Karel Spruyt for figures and photographs. This work was supported by grants from the Services of the Prime Minister (IUAP I2OCUI92) and the Commission of the European Communities (TSD-A-0135B and BAP-0173-DK). M. H. is a research associate of the National fund for Scientific Research, Belgium. P. M. thanks the Instituut ter aanmoediging van het Wetenschappelijk Onderzoek in de Nijverheid en de Landbouw (IWONL) for a predoctoral fellowship.

#### LITERATURE CITED

Baev, N., Endre, G., Petrovics, G., Banfalvi, Z., and Kondorosi, A. 1991. Six nodulation genes of *nod* box locus 4 in *Rhizobium meliloti* are involved in nodulation signal production: *nodM* codes for D-

- glucosamine synthetase. *Mol. Gen. Genet.* 228:113-124.
- Bolivar, F. 1978. Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique *EcoRI* sites for selection of *EcoRI* generated recombinant DNA molecules. *Gene* 4:121-136.
- Casadaban, M. J., and Cohen, S. N. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* 138:179-207.
- Dénarié, J., and Roche, P. 1991. *Rhizobium* nodulation signals. Pages 295-324 in: *Molecular Signals in Plant-Microbe Communications*. D. P. S. Verma, ed. CRC Press, Boca Raton, FL.
- Ditta, G., Stanfield, S., Corbin, D., and Helinski, D.R. 1980. Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* 77:7347-7351.
- Dreyfus, B., Garcia, J. L., and Gillis, M. 1988. Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., a stem-nodulating nitrogen-fixing bacterium isolated from *Sesbania rostrata*. *Int. J. Syst. Bacteriol.* 38:89-98.
- Dreyfus, B., and Dommergues, Y. R. 1981. Nitrogen-fixing nodules induced by *Rhizobium* on the stem of the tropical legume *Sesbania rostrata*. *FEMS Microbiol. Lett.* 10:313-317.
- Friedman, A. M., Long, S. R., Brown, S. E., Buikema, W. J., and Ausubel, F. M. 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. *Gene* 18:289-296.
- Goethals, K., Gao, M., Tomekpe, K., Van Montagu, M., and Holsters, M. 1989. Common *nodABC* genes in Nod locus 1 of *Azorhizobium caulinodans*: Nucleotide sequence and plant-inducible expression. *Mol. Gen. Genet.* 219:289-298.
- Goethals, K., Van den Eede, G., Van Montagu, M., and Holsters, M. 1990. Identification and characterization of a functional *nodD* gene in *Azorhizobium caulinodans* strain ORS571. *J. Bacteriol.* 172:2658-2666.
- Goethals, K., Van Montagu, M., and Holsters, M. 1992. Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proc. Natl. Acad. Sci. USA* 89:1646-1650.
- Heffron, F., Bedinger, P., Champoux, J. J., and Falkow, S. 1977. Deletions affecting the transposition of an antibiotic resistance gene. *Proc. Natl. Acad. Sci. USA* 74:702-706.
- Krishnan, H. B., and Pueppke, S. G. 1991. Repetitive sequences with homology to *Bradyrhizobium japonicum* DNA and the T-DNA of *Agrobacterium rhizogenes* are closely linked to *nodABC* of *Rhizobium fredii* USDA257. *Mol. Plant-Microbe Interact.* 4:521-529.
- Lemaire, H.-G., and Müller-Hill, B. 1986. Nucleotide sequences of the *galE* gene and the *galT* gene of *E. coli*. *Nucleic Acids Res.* 14:7705-7711.
- Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J. C., and Dénarié, J. 1990. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acetylated glucosamine oligosaccharide signal. *Nature* 344:781-784.
- Long, S. R. 1992. Genetic analysis of *Rhizobium* nodulation. Pages 560-597 in: *Biological Nitrogen Fixation*. G. Stacey, R. H. Burris, and H. J. Evans, eds. Chapman & Hall, New York.
- Martinez, E., Romero, D., and Palacios, R. 1990. The *Rhizobium* genome. *Crit. Rev. Plant Sci.* 9:59-93.
- Messens, E., Geelen, D., Van Montagu, M., and Holsters, M. 1991. 7,4'-Dihydroxyflavanone is the major *Azorhizobium nod* gene-inducing factor present in *Sesbania rostrata* seedling exudate. *Mol. Plant-Microbe Interact.* 4:262-267.
- Ndoye, I. 1990. Contribution à l'étude de la symbiose entre *Azorhizobium*, *Rhizobium* et *Sesbania rostrata*. Thèse de doctorat d'Université. Lille (France) Université des Sciences et Techniques.
- Pissowotzki, K., Mansouri, K., and Piepersberg, W. 1991. Genetics of streptomycin production in *Streptomyces griseus*: Molecular structure and putative function of genes *strELMB2N*. *Mol. Gen. Genet.* 231:113-123.
- Prentki, P., and Krisch, H. M. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. *Gene* 29:303-313.
- Rossmann, M. G., Liljas, A., Brenden, C. I., and Banaszak, L.J. 1975. Evolutionary and structural relationships among dehydrogenases. Pages 61-102 in: *The Enzymes*, 3rd ed. P. D. Boyer, ed. Academic Press, New York.
- Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A., and Kondorosi, A. 1986. Conservation of extended promoter regions of nodulation genes in *Rhizobium*. *Proc. Natl. Acad. Sci. USA* 83:1757-1761.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Schwedock, J., and Long, S. R. 1990. ATP sulphurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. *Nature* 348:644-647.
- Spaink, H. P., Okker, R. J. H., Wijffelman, C. A., Pees, E., and Lugtenberg, B. J. J. 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRLIJ1. *Plant Mol. Biol.* 9:27-39.
- Spaink, H. P., Sheeley, D. M., van Brussel, A. A. N., Glushka, J., York, W. S., Tak, T., Geiger, O., Kennedy, E. P., Reinhold, V. N., and Lugtenberg, B. J. J. 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* 354:125-130.
- Stormo, G. D., Schneider, T. D., and Gold, L. M. 1982. Characterization of translational initiation sites in *E. coli*. *Nucleic Acids Res.* 10:2971-2996.
- Szikora, J.-P., Van Pel, A., Brichard, V., André, M., Van Baren, N., Henry, P., De Plaen, E., and Boon, T. 1990. Structure of the gene of tum<sup>-</sup> transplantation antigen P35B: Presence of a point mutation in the antigenic allele. *EMBO J.* 9:1041-1050.
- Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., de Billy, F., Promé, J.-C., and Dénarié, J. 1991. Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. *Nature* 351:672-673.
- Van den Eede, G., Dreyfus, B., Goethals, K., Van Montagu, M., and Holsters, M. 1987. Identification and cloning of nodulation genes from the stem-nodulating bacterium ORS571. *Mol. Gen. Genet.* 206:291-299.
- Wheatcroft, R., and Laberge, S. 1991. Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence IS*Rm3*: Similarity between the putative transposase encoded by IS*Rm3* and those encoded by *Staphylococcus aureus* IS256 and *Thiobacillus ferrooxidans* IST2. *J. Bacteriol.* 173:2530-2538.