JOURNAL OF BACTERIOLOGY, Feb. 1991, p. 1250–1258
0021-9193/91/031250-09$02.00/0

Vol. 173, No. 3

Novel Organization of the Common Nodulation Genes in *Rhizobium leguminosarum* bv. phaseoli Strains

MARTHA VÁZQUEZ,* ARACELI DÁVALOS, ALEJANDRO DE LAS PEÑAS, FEDERICO SÁNCHEZ, AND CARMEN QUINTO

Unidad de Biología Molecular y Biotecnologia Vegetal, Centro de Investigación sobre Fijación de Nitrógeno, UNAM, Apartado Postal 2-246, Cuernavaca, Morelos, Mexico

Received 4 October 1990/Accepted 3 December 1990

Nodulation by *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* species in the roots of legumes and nonlegumes requires the proper expression of plant genes and of both common and specific bacterial nodulation genes. The common *nodABC* genes form an operon or are physically mapped together in all species studied thus far. *Rhizobium leguminosarum* bv. phaseoli strains are classified in two groups. The type I group has reiterated *nifHDK* genes and a narrow host range of nodulation. The type II group has a single copy of the *nifHDK* genes and a wide host range of nodulation. We have found by genetic and nucleotide sequence analysis that in type I strain CE-3, the functional common *nodA* gene is separated from the *nodBC* genes by 20 kb and thus is transcriptionally separated from the latter genes. This novel organization could be the result of a complex rearrangement, as we found zones of identity between the two separated *nodA* and *nodBC* regions. Moreover, this novel organization of the common *nodABC* genes seems to be a general characteristic of *R. leguminosarum* bv. phaseoli type I strains. Despite the separation, the coordination of the expression of these genes seems not to be altered.

The common nodulation genes *nodA*, *nodB*, and *nodC* have been found in different *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* species and are required for the initiation of nodulation of legumes and nonlegumes. The *nodABC* genes have always been found together in all species studied so far. In some species, the *nodABC* genes form an operon, as shown by genetic analysis (2, 29, 39). In others, there is indirect evidence, such as nucleotide sequence analysis (15, 45, 46, 50) or induction of *nodABC*::*lacZ* fusions in the presence of flavonoids (2, 15, 16, 23), suggesting that the common *nod* genes could be transcribed as a polycistronic unit. In a few other cases, these genes have only been physically mapped together (1, 34).

Mutations in the *nodA*, *nodB*, or *nodC* gene prevent early events in the infection process, such as root hair curling and the initial cortical cell divisions in the root tissue (for a review, see reference 37), and completely abolish the ability to nodulate. Recent studies on the function of the Nod proteins suggest that the NodA and NodB proteins are involved in generating small, heat-stable compounds that stimulate the mitosis of various legume and nonlegume protoplasts (44). The NodA protein is found in the cytoplasm and cell envelope (21), and the NodB protein is also located in the cytosol (44). NodC is a cell surface protein with a eukaryotic receptor-like structure, which may serve as a transducer of an intracellular bacterial signal to root cells (20). It has been proposed (11) that the common *nodABC* genes are responsible for production of a basic, common signal which could be modified by the action of the host range genes. This specific factor, NodRm1, has recently been purified in *Rhizobium meliloti*; it is a sulfated β-1,4-tetrasaccharide of D-glucosamine (24).

The product of the *nodD* gene, acting as a positive regulator in combination with plant flavonoid compounds, induces the expression of common *nodABC* genes and other *nod* genes. The NodD product has been found in the cytoplasmic membrane, and its localization does not change in the presence of the inducer (43).

The NodD protein binds to the *nod* box (13, 14, 17, 22), a strongly conserved regulatory sequence found upstream of all inducible *nod* genes (40). Negative regulation of *nod* gene expression has been reported for some genetic backgrounds in *R. meliloti* and is mediated by a putative repressor recognizing the RNA polymerase binding site (22). The mechanism of interaction between NodD, the *nod* box, the flavonoids, the repressor (if such is the case), and RNA polymerase to promote expression is not yet well established.

Many flavonoid compounds produced by seeds or root exudates have been described as having particular capacities to induce or repress transcription of the common *nodABC* genes and other *nod* genes (reviewed in reference 36). This specificity depends on the presence of the different *nodD* products described for each species. In contrast to other fast-growing rhizobia, it has been shown that in the *R. leguminosarum* bv. phaseoli type I strain CE-3, isoflavones such as genistein and diadzein are good inducers of common *nod* gene expression, as is the flavanone naringenin (41).

The *R. leguminosarum* bv. phaseoli group has been classified into two types. The type I strains are defined by the reiteration of the *nifHDK* genes and a narrow host range of nodulation. In contrast, the type II strains have one copy of the *nifHDK* genes and have a broad host range of nodulation (4).

Studies of *R. leguminosarum* bv. phaseoli 8002 (8, 9) demonstrate that it harbors three functional *nodD* copies, two of them tightly linked and the third mapping near a common *nod* region, defined by hybridization to a *nodABC* probe from *R. leguminosarum* bv. viciae. Nevertheless, it is not clear from those studies whether the *nodA* gene is present in this hybridizing region.

* Corresponding author.

TABLE 1. Bacterial strains, plasmids, cosmids, and bacteriophages used

Strain, plasmid, cosmid, or bacteriophage	Relevant characteristics	Source or reference
Strains		
R. leguminosarum bv. phaseoli		
CE-3	Spontaneous Str^r derivative of wild-type strain CFN42; Nod^+ Fix^+	31
CFN-2001	Rif^r derivative of CFN-42 cured of plasmids p42a and p42d; Nod^- Fix^-	32
UBP101	CE-3 with pRp30	This work
UBP102	CE-3 derivative with *nodA*::Mu dIIPR13 in p42d	This work
UBP201	CFN-2001 with pRp30	This work
UBP301	CE-3 with pRp32	This work
UBP401	CFN-2001 with pRp32	This work
Bra-8	*R. leguminosarum* bv. phaseoli type I strain	33
Viking-I	*R. leguminosarum* bv. phaseoli type I strain	27
Nitragin-8251	*R. leguminosarum* bv. phaseoli type I strain	27
CFN-3	*R. leguminosarum* bv. phaseoli type I strain	27
E. coli		
MC4100 (Mu cts)	*araD139* Δ(*argF-lacIPOZYA*)*U169 rpsL150 relA1 flbB5301 ptsF25 deoC1* (Mu cts)	6
M8820 (Mu c^+)	*araD139* Δ(*araCOIBA-leu*)*7679* Δ(*proAB-argF-lacIPOZYA*)*XIII rpsL* (Mu c^+)	5
HB101	F^- *hsdS20-recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 supE44*	3
Plasmids		
pRK2013	ColE1 *mob*$^+$ Tra^+ (RK2) Km^r	12
pNC206	IncP1 Cb^r Km^r	A. Pühler
pUC19	ColE1 replicon; Ap^r	53
Cosmids		
pSUP205	ColE1 replicon; *cos* Tc^r Cm^r	47
pSM991.25	pSUP205 with 18.5 kb from p42d cloned in the *Eco*RI site; Tc^r Cm^s	7
pSM991.44	pSUP205 with a 3.5-kb *Eco*RI fragment containing *nodA* from pSM991.25	7
pRp30	pSM991.44 with *nodA*::Mu dIIPR13; Tc^r Cm^r	This work
pRp32	pSUP205 with a 6.8-kb *Eco*RI fragment from pSM991.25 *nodC*::Mu dIIPR13	This work
Bacteriophage		
Mu dIIPR13	*lac'ZYA* Cm^r	35

In a previous work (7), we identified two nodulation regions from the symbiotic plasmid of *R. leguminosarum* bv. phaseoli type I strain CE-3 by complementation of a nodulation-deficient strain and by insertional mutagenesis analysis. Mutations in both regions prevent nodulation of bean roots by this complemented strain. Region I, a 6.8-kb *Eco*RI fragment, was associated with the common *nod* genes by heterologous DNA-DNA hybridization criteria, using *nodAB*, *nodBC*, and *nodC R. meliloti* probes. Region II, a 3.5-kb *Eco*RI fragment, did not hybridize with these *R. meliloti* common *nod* gene probes in those experiments (7). Here, we show that region II does contain the functional and essential common *nodA* gene and that the separation of *nodA* from *nodBC* is a common arrangement in *R. leguminosarum* bv. phaseoli type I strains.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. Bacterial strains and plasmids are described in Table 1. *R. leguminosarum* bv. phaseoli strains were grown in peptone-yeast medium (31) or in medium containing yeast extract and mannitol (18) for determination of β-galactosidase activities. *Escherichia coli* strains were grown as described by Maniatis et al. (26) or in Luria broth medium (28) supplemented with 5 mM $CaCl_2$ and 200 mM $MgSO_4$. The media were supplemented with tetracycline (3 μg/ml), chloramphenicol (12 μg/ml), nalidixic acid (20 μg/ml), kanamycin (30 μg/ml), streptomycin (100 μg/ml), and rifampin (50 μg/ml) for *Rhizobium* strains and with tetracycline (10 μg/ml), ampicillin (200 μg/ml), chloramphenicol (30 μg/ml), and kanamycin (30 μg/ml) for *E. coli* strains. *Rhizobium* strains and *E. coli* thermosensitive strains were grown at 30°C. Other *E. coli* strains were grown at 37°C. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and isopropylthiogalactoside (IPTG) were added to the media at a final concentration of 25 μg/ml each when needed.

Genetic techniques. To obtain *nodA*::*lacZ* translational fusions, cosmid pSM991.44 (Tc^r), which carries the *R. leguminosarum* bv. phaseoli CE-3 *nodA* gene, was mutagenized with the bacteriophage Mu dIIPR13, which has the *E. coli lacZYA* operon without a promoter and the *cat* gene (Cm^r) (35). The cosmid was introduced in *E. coli* MC4100(Mu cts)(Mu dIIPR13) by triparental mating, using pRK2013 (Km^r) as a helper plasmid and selecting for Cm^r, Tc^r, and Km^s colonies. Thermoinduction of the bacteriophage was performed as described previously (7). After infection, Cm^r and/or Tc^r transductants were selected.

Homogenotization was done by triparental mating of the CE-3 wild-type receptor strain, using pRK2013 as a helper plasmid and M8820(Mu c^+)/pRp30 (Cm^r Tc^r) and selecting for Cm^r Tc^s colonies.

DNA manipulation and sequencing. Recombinant DNA techniques were carried out as described by Maniatis et al. (26). DNA sequencing was done by the dideoxy-chain termination method (42), using the pUC vector system (53) and the Sequenase version 2.0 sequencing kit (U.S. Biochemical Corp.). To sequence the 5'-3' chain of the *nod* box region and the 3'-5' chain from nucleotides 45 to 62 from the *nodA* gene, two oligonucleotides of 20- and 22-base length with the sequences 5'-GAAGTTGGAGCCCGGCCGCT-3' and 5'-GGGCGTTGTAAGCTCCAGTTGG-3', respectively, were used. The nucleotide sequence reported was read in both DNA strands at least twice, and all overlaps were checked except nucleotides −290 to −196 in Fig. 2, for which only the 5'-3' chain was read.

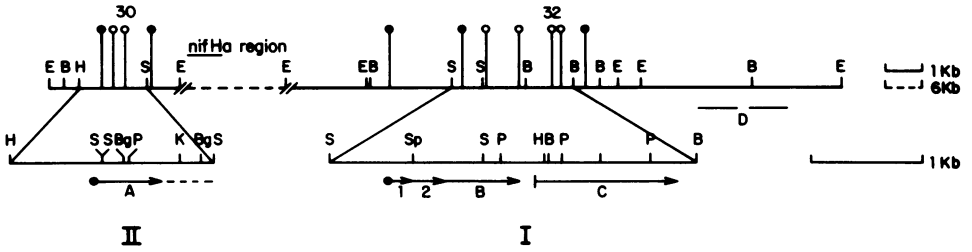

FIG. 1. Nodulation region of symbiotic plasmid p42d from strain CE-3 harbored in pSM991.25. Sequenced regions are amplified. E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; S, *Sst*I; P, *Pst*I; Bg, *Bgl*II; K, *Kpn*I; A, *nodA*; 1 and 2, ORF1 and ORF2, respectively; B, *nodB*; C, *nodC*; D, *nodD*; I, region I; II, region II. The broken line downstream of *nodA* indicates the *nodB*-like region; vertical lines indicate different insertions sites isolated in pSM991.25; *nod* boxes are indicated by filled circles at the left of the indicated ORFs or genes; 30 and 32 indicate the insertion sites of Mu dIIPR13 in pRp30 and pRp32, respectively. ●, Nod^+ phenotype; ○, Nod^- phenotype of strain CFN-2001 harboring pSM991.25 with different insertions. The broken line in the nonamplified map indicates a deletion of the original isolate.

The *R. leguminosarum* bv. phaseoli gene probes used in DNA-DNA hybridizations of total DNAs of different type I strains were a 0.3-kb *Pst*I-*Sst*I fragment from *nodA*, a 0.7-kb *Sph*I-*Pst*I fragment from *nodB*, and a 0.9-kb *Bam*HI-*Pst*I fragment from *nodC*. All probes are intragenic except for the *nodB* probe, which also carries almost the entire open reading frame 2 (ORF2) region.

Homologous DNA-DNA hybridizations were done in phosphate buffer at 65°C, and heterologous hybridizations were done in 30% formamide at 42°C (26).

Computer sequence analysis. The nucleotide sequence analyses were carried out by using the Genetics Computer Group sequence analysis software package (version 6.0) (10). Initially, the analysis was made with the FASTA and FASTAP programs. The percentages of nucleotide and peptide identities and similarities were obtained from the GAP program.

Determination of β-galactosidase activities. Assays of β-galactosidase were carried out as described by Zaat et al. (54). Naringenin (120 nM), genistein (440 nM), or bean root exudate was added, depending on the case. The cultures were concentrated 10-fold in 10 mM $MgSO_4$; 0.1 ml was used for β-galactosidase activity determination (28), and 0.1 ml was used for determination of protein concentration (25).

Preparation of bean root exudates. The procedure was done as described by van Brussel et al. (51) and Zaat et al. (54). Sterile exudates were concentrated 10-fold by vacuum evaporation at 0°C and were used as 10-fold stocks in induction growth conditions for the determination of β-galactosidase activities. As for naringenin and genistein, higher concentrations of exudate do not cause higher induction.

Plant nodulation tests. The nodulation tests were done as described before (7). Bacteria from nodules were isolated and checked for antibiotic markers and for the DNA-DNA hybridizations patterns.

Nucleotide sequence accession numbers. The sequences reported for regions I and II will appear in the EMBL, GenBank, and DDBJ nucleotide sequence data bases under accession numbers M58626 and M58625, respectively.

RESULTS

Genetic analysis of region II. As shown in our previous work (7), cosmid pSM991.25 (Fig. 1) is a deletion derivative of cosmid pSM991 carrying 18 kb from the strain CE-3 symbiotic plasmid (pSym). pSM991.25 has two essential regions for nodulation when tested in a $pSym^-$ background (strain CFN-2001): region I, a 6.8-kb *Eco*RI fragment that hybridizes with *R. meliloti nodAB*, *nodBC*, and *nodC* probes, and region II, a 3.5-kb *Eco*RI fragment that does not hybridize with any of these probes. To test whether the information harbored in region II was also essential for nodulation in the wild-type background, we homogenotized a mutation in region II in the strain CE-3 symbiotic plasmid. For that purpose, we constructed a Mu dIIPR13 insertion collection in cosmid pSM991.44 which harbors only region II, as described in Materials and Methods. A total of 400 mutations were screened for inducibility by 120 nM naringenin in MM media with X-Gal (54); the insertions in 30 inducible colonies were mapped with the restriction enzymes *Eco*RI, *Bam*HI, and *Pst*I. We selected cosmid pRp30 for homogenotization because it had an insertion mapping exactly at the same site (Fig. 1) of the previously obtained mutations in cosmid pSM991.25 (pRp23 and pRp24) (7).

The mutation contained in pRp30 was homogenotized as described in Materials and Methods. Various colonies were screened by Southern hybridization, using as a probe cosmid pSM991.44, to verify that the mutation had recombined with its homologous wild-type sequence (data not shown). The resulting strain was designated UBP102. Five independent colonies of this strain were used to inoculate five bean seedlings each, to test their nodulation ability. None of the inoculated plants had any nodules after 20 days of growth.

We complemented strain UBP102 with cosmid pSM991.44, which carries the wild-type allele in the 3.5-kb *Eco*RI fragment. Three colonies from the complemented strain were tested for bean nodulation. All plants inoculated with this strain had nodules at the same time, number, and size as the plants inoculated with strain CE-3.

Analysis of the nucleotide sequences of region I and II genes of pSM991.25. The information obtained from the nucleotide sequences of regions I and II is presented in Fig. 1 to 3.

In region II (Fig. 2), we found a 585-nucleotide-long ORF. This ORF codifies for a putative peptide with a molecular weight of 21,870 and shows identity to the NodA peptide sequence from other species (Table 2). These data, taken together with the Nod^- phenotype of strain UBP102, lead us to propose that this ORF encodes the *R. leguminosarum* bv. phaseoli CE-3 *nodA* gene. We sequenced approximately 400 nucleotides downstream of the *nodA* gene. The longer ORFs found in the 5′-3′ and in the 3′-5′ strands of this region are 43 and 61 amino acids long, respectively, and they are still open at the end of the reported sequence. Searches of data bases for homology of this 400-nucleotide sequence revealed no significant identity.

Region I has four ORFs (Fig. 3). ORF3 spans from nucleotides 455 to 1141. The ORF3 derived peptide sequence

FIG. 2. Nucleotide sequence of region II. Region II is indicated in Fig. 1. The predicted peptide sequence of *nodA* and the relevant restriction sites are shown. ORF2 and *nodB*-like sequences are indicated with a line over the nucleotide sequence.

has identity to the NodB peptide sequence from other species (Table 2). The calculated molecular weight (25,061) is also similar to the NodB molecular weights in these strains. We propose that ORF3 is the *nodB* gene in strain CE-3. The ATG codon of *nodB* overlaps the termination codon of ORF2.

ORF4 has three putative translation initiation codons, at nucleotides 1249, 1255, and 1321, and ends at nucleotide 2532. The ORF4 peptide has sequence identity to the NodC products from different strains (Table 2). The predicted molecular weight for the ORF4 product is 47,236 (using the ATG codon at position 1249) or 44,685 (using the third ATG codon at nucleotide 1321). The calculated molecular weight of 44,685 is also similar to those of the NodC products from the mentioned strains (Table 2). These data lead us to propose that ORF4 is the *nodC* gene in strain CE-3. The distance between the *nodB* stop codon and the initiation of the *nodC* gene is 177 nucleotides if the *nodC* ATG codon is at position 1321 and 104 or 110 nucleotides if the ATG codon is at nucleotide 1249 or 1255, respectively.

ORF1 spans 171 nucleotides, from nucleotides 0 to 170, and has no similarity to any gene included in the GenBank and EMBL data banks. It has two putative Shine-Dalgarno sequences, but one of them is not at the correct position according to the criteria proposed by Stormo et al. (49) and the other is not identical to the consensus Shine-Dalgarno sequence. The putative ORF1 product molecular weight is 6,279. We do not know whether this ORF is translated.

ORF2 has 273 bases, from nucleotides 183 to 455, and encodes a possible product with a molecular weight of 10,476. It has a Shine-Dalgarno sequence at the correct position, but we do not know whether it encodes a functional product. In other species, the *nodA* stop codon overlaps the *nodB* initiation codon. This ORF ends with the sequence ATGA, and thus the putative ORF2 stop codon overlaps the *nodB* initiation codon. ORF2 has two zones of identity to the *nodA* gene, located 20 kb away (Fig. 4A). We did not detect any identity of the first region found almost at the ORF2 5′ end to other *nodA* genes from other species. In contrast, a second 119-nucleotide-long *nodA*-like region beginning at

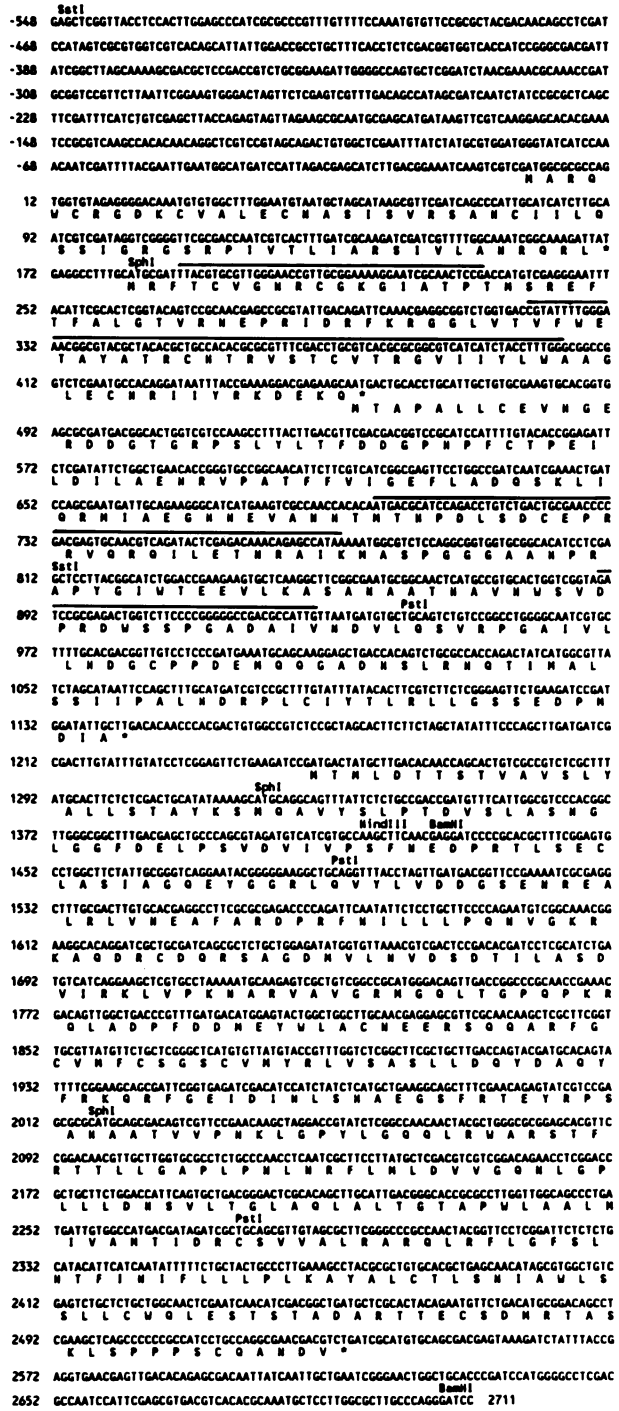

FIG. 3. Nucleotide sequence of region I. Region I, contained in an *Sst*I-*Bam*HI fragment, is shown in Fig. 1. The predicted peptide sequences of ORF1, ORF2, *nodB*, and *nodC* are shown, as are the relevant restriction sites. *nodA*-like regions in ORF2 and the *nodB* regions identical to sequences downstream of *nodA* are indicated with a line over the nucleotide sequence.

the middle of ORF2 (Fig. 2, 3, and 4A) also has identity to the *nodA* genes from other species, and in some cases this identity is extended, as in the *R. leguminosarum* bv. viciae *nodA* gene (GenBank number Y00548, 1989), where there are 133 nucleotides 64% identical.

TABLE 2. NodABC identities and similarities between different species

Strain[a]	Determination[b]					
	NodA		NodB		NodC[c]	
	I	S	I	S	I	S
Rhizobium leguminosarum bv. viciae	64	80	59	75	55	70
R. leguminosarum bv. trifolii	64	80	54	69	—[d]	—
R. meliloti 1021	60	76	59	71	56	67
R. meliloti 41	61	76	60	74	56	67
Bradyrhizobium parasponia	65	81	58	72	—	—
Azorhizobium caulinodans	54	69	36	54	42	61

[a] The NodABC peptide sequences were obtained from GenBank: Y00548, X03721, M11268, and Jacobs et al. (19); X01649, X03720, and Goethals et al. (15).
[b] Percentages of identity (I) and similarity (S) to the indicated CE-3 gene, obtained as described in Materials and Methods.
[c] The comparison was done by using the initiation codon at nucleotide 1249.
[d] —, Not determined.

Downstream of the *nodA* gene, we found *nodB*-like sequences (Fig. 2, 3, and 4B). The 95-nucleotide first zone shown, corresponding to the middle of *nodB*, also has identity to the same regions of the *nodB* nucleotide sequences from other species, the largest (135 nucleotides with 60% identity) being in the *nodB* gene nucleotide sequence from *R. meliloti* 41 (50). There are some other shorter regions of identity with the *nodB* genes, one of which is shown in Fig. 4B.

The *nodA* gene is preceded by a sequence spanning from nucleotides −146 to −75 that is 85% similar to the consensus *nod* box sequence proposed by Spaink et al. (48) (Fig. 4C). The end of this *nod* box sequence is 74 nucleotides away from the proposed *nodA* ATG codon. We found a sequence

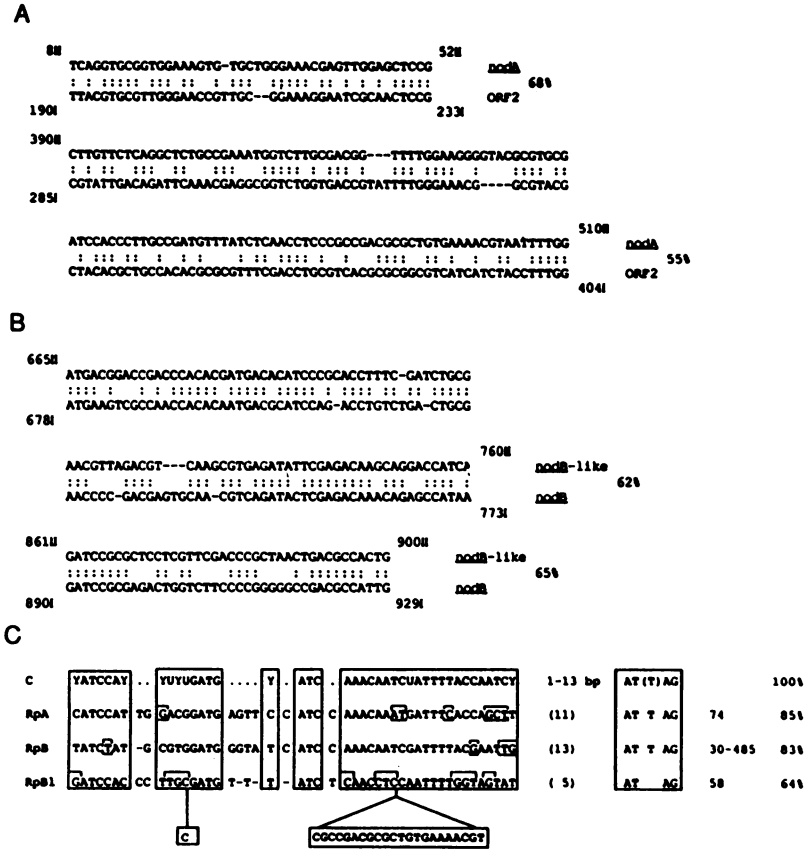

FIG. 4. Nucleotide sequence identities and *nod* boxes in regions I and II. (A) Identities between *nodA* and ORF2. (B) Identities between *nodB* and the *nodB*-like sequence. To the left and to the right of the sequences are indicated the first and last nucleotides delimiting these zones, which are also indicated in Fig. 1. (C) *nod* boxes found in regions I and II. C, Consensus *nod* box sequence (48); RpA, *nodA nod* box; RpB, ORF1-ORF2-*nodB*-*nodC nod* box; RpB1, pseudo-*nod* box from region II preceding the *nodB*-like sequences. Regions from the consensus sequence well conserved between different *nod* boxes are boxed; bases different from the consensus are outside the boxed sequences. Numbers to the right of the *nod* boxes indicate the distance to the first ATG found. In the case of RpB, 30 and 485 are the distances to ORF1 and to *nodB*, respectively. Also shown are the identity percentages.

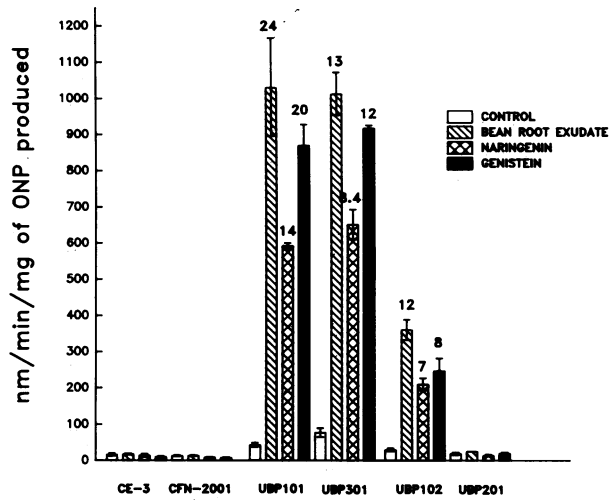

FIG. 5. Expression of *nodA*::*lacZ* and *nodC*::*lacZ* fusions in *R. leguminosarum* bv. phaseoli strains. All strains are described in Table 1. β-Galactosidase activities were measured at least three times, and all standard deviations are less than 20%. Cultures were grown without inducer, with naringenin (120 nM), with genistein (440 nM), or with bean root exudate as described in Materials and Methods. Numbers above the bars indicate fold induction with respect to the control. ONP, *o*-Nitrophenol.

that is 64% identical to the *nod* box consensus sequence but with a 22-base-long insertion (Fig. 4C) in the 3′ end of the *nodA* coding region, from nucleotides 458 to 525. This sequence is 58 nucleotides upstream of the TGA codon of the *nodA* gene preceding the mentioned *nodB*-like sequences. The ORF1-ORF2-*nodB*-*nodC* region is also preceded by a sequence from nucleotides −95 to −31 which is 83% identical to the consensus *nod* box sequence (Fig. 4C). There are no sequences suggestive of promoters or terminators between the *nod* box sequence and the end of *nodC*.

Expression of genes found in regions I and II of cosmid pSM991.25. We measured the levels of expression of *nodA*::*lacZ* and *nodC*::*lacZ* fusions (pRp30 and pRp32, respectively; Table 1 and Fig. 1) in the presence (strains UBP101 and UBP301) or absence (strains UBP201 and UBP401) of the symbiotic plasmid. We also measured expression of the *nodA* homogenotized mutation in the symbiotic plasmid in strain UBP102. The flavanone naringenin, the isoflavone genistein, and bean root exudates were used as inducers of expression.

When plasmid pRp30 or pRp32 was in strain CFN-2001 (pSym cured), there was no induction in any case (Fig. 5; data not shown for UBP401). When these plasmids were in the CE-3 background (pSym$^+$), naringenin induced β-galactosidase expression from a basal level of 43 to 592 nm/min/mg in pRp30 and from 77 to 651 nm/min/mg in pRp32, representing increments of 14- and 8.4-fold, respectively.

Genistein was a slightly better inducer, although a 440 nM concentration was required to obtain maximal induction. Strains harboring plasmid pRp30 had activity of 870 nm/min/mg, and those with pRp32 had activity of 917 nm/min/mg, which represented 20- and 12-fold induction, respectively. Strain UBP102, which carries the *nodA* homogenotized mutation, maintained the same regulation pattern as the multicopy plasmid pRp30, although since this strain carries the mutation in only one copy, the absolute values of β-galactosidase activities were lower. The basal level of β-galactosidase activity was 30 nm/min/mg; naringenin induced expression to 210 nm/min/mg (sevenfold induction), and genistein increased expression to 247 nm/min/mg (eightfold induction).

In all of the mutant pSym$^+$ strains tested, the bean root exudates were the best inducers, giving activity levels of 1,031 nm/min/mg for strain UBP101 (pRp30 plasmid), 1,013 nm/min/mg for strain UBP301 (pRp32 plasmid), and 361 nm/min/mg for strain UBP102, representing 24-, 13-, and 12-fold induction, respectively.

Thus, the fold induction of the *nodC*::*lacZ* fusion was, in general, almost half that of the *nodA*::*lacZ* fusion with all inducers tested, when measured in the multicopy plasmids harbored in the wild-type background. In the pSym-cured background there was no induction of any of these fusions in any condition.

Separation of *nodA* from *nodBC* genes is a common feature of *R. leguminosarum* bv. phaseoli type I strains. We have shown that in strain CE-3, *nodA* is separated from *nodBC* genes by 20 kb. To test whether this arrangement is common to other *R. leguminosarum* bv. phaseoli type I strains, we made a restriction fragment length polymorphism (RFLP) analysis of *nodA*, *nodB*, and *nodC* genes in other strains. We selected five strains representing the genetic diversity found by Piñero et al. (33) in a population genetics study of this group. Total DNAs from strains Viking-I, Nitragin-8251, CFN-3, Bra-8, and CE-3 were digested with *Eco*RI, *Hin*dIII, and *Sst*I and were hybridized with *nodA*-, *nodB*-, and *nodC*-specific probes obtained from strain CE-3. The RFLP patterns obtained with the *nodB* and *nodC* probes had the same signal in two of the three restriction enzyme pattern (Table 3), which strongly suggests that in all of these strains *nodB* and *nodC* are found together. In contrast, the RFLP pattern obtained with the *nodA* probe never coincided with the *nodB* and *nodC* RFLP patterns (Table 3), indicating that in all of these strains the *nodA* gene is separated from *nodB* and *nodC* genes.

DISCUSSION

We report here that in *R. leguminosarum* bv. phaseoli CE-3, the unique and functional *nodA* gene is separated by 20 kb from the *nodBC* genes. This arrangement apparently is a general feature of the *R. leguminosarum* bv. phaseoli type I strains.

In a previous work (7), we found that the cosmid pSM991.25 harbors two essential regions for nodulation; region I showed similarity to plasmid pKSK5, which harbors the *nodABC* genes, but we were not able to find any similarity to region II with this plasmid. Further experiments have shown that the identity does exist, and this conclusion was finally confirmed by nucleotide sequence data analysis (Fig. 2 and 3).

Strain UBP102 harbors a homogenotized *nodA* mutation in pSym. This strain had a Nod$^-$ phenotype. When we complemented strain UBP102 with its respective wild-type region in *trans*, the Nod$^+$ phenotype was restored, demonstrating that the altered nodulation phenotype is caused only by the introduced mutation. On the other hand, a homogenotized *nodC* insertion in pSym and mutations in the *nodB* gene harbored in cosmid pSM991.25 in strain CFN-2001 also cause a Nod$^-$ phenotype (7). Thus, these *nodABC* genes are unique and essential for nodulation in strain CE-3.

Nucleotide sequence analysis of the genes found in both regions describes a completely different organization of common *nod* genes in strain CE-3 (Fig. 1) compared with

TABLE 3. *nodA*, *nodB*, and *nodC* RFLPs of different *R. leguminosarum* bv. phaseoli type I strains

Strain[a]	Hybridizing fragment size (kb)[b]								
	*Eco*RI			*Hin*dIII			*Sst*I		
	nodA	*nodB*	*nodC*	*nodA*	*nodB*	*nodC*	*nodA*	*nodB*	*nodC*
CE-3	3.5	6.9	6.9	4.8	9.1	15.1	1.0	5.0 1.4	4.9
Bra-8	3.6	4.2	4.2	5.2	2.4	15.1	1.0	5.6	5.6
Viking-I	4.7	5.6	5.6	6.4	12.6	15.1	1.0	5.6	5.6
Nitragin-8251	3.8	7.2	7.2	5.4	8.7	15.1	1.0	5.0 1.4	4.9
CFN-3	3.5	5.7	5.9	3.5	8.5	15.1	1.0	5.6	5.6

[a] See Table 1 for descriptions.

[b] Determined by Southern blot hybridization of total DNA digested with *Eco*RI, *Hin*dIII, and *Sst*I and probed with *nodA*, *nodB*, and *nodC R. leguminosarum* bv. phaseoli CE-3 probes. Underlined numbers indicate that the same *Eco*RI or *Sst*I fragments hybridized when the *nodB* or *nodC* probe was used in each strain.

that reported for all of the strains from different genera described up to now, including a strain from the *R. leguminosarum* bv. phaseoli type II group (52). Nevertheless, the molecular weights and derived peptide sequences of the products from these genes are very similar to those found for the NodABC products from other species (Table 2).

Upstream of the *nodB* and *nodC* genes are two very short ORFs. Both ORFs are preceded by a *nod* box and have putative Shine-Dalgarno sequences, but we do not know whether they are translated. The mutations in this region conferred a Nod$^-$ phenotype when harbored in cosmid pSM991.25 in a pSym$^-$ background, but this phenotype could be due to a polarity over the genes localized downstream these two ORFs. ORF1 has no identity or similarity to any *nod* gene reported. Further analysis must be done to establish the functionality of these sequences.

In other species in which the *nodABC* genes form an operon, the *nodA* translation stop codon overlaps the *nodB* initiation codon, as found here for the proposed ORF2 stop codon, which overlaps the *nodB* initiation codon.

In strain CE-3, the distance from the *nodB* stop codon to any of the putative ATG codons of *nodC* is more than 100 nucleotides. In contrast, this distance is 20 to 30 nucleotides in other rhizobia (15, 38, 45, 46, 50). There are no putative promoters or terminators in this long sequence. It seems very likely that ORF1, ORF2, *nodB*, and *nodC* could be transcribed as an operon.

There is some identity between nucleotide sequences localized in regions I and II (Fig. 4A and B). These identities are between ORF2 and the *nodA* gene and between a sequence localized downstream of the *nodA* and *nodB* genes. Moreover, these identities are also found, and in some cases are more extended, with the common *nod* genes from other species; for instance, ORF2 has the most extended identity (64%) to the *nodA* gene from *R. leguminosarum* bv. viciae (GenBank number Y00548), while the most extended identity found from the *nodB*-like sequences, downstream of the *nodA* gene, is to the *R. meliloti* 41 *nodB* gene (50). We sequenced only 400 nucleotides downstream of the *nodA* gene, and we do not know whether the identity to *nodB* extends beyond this point. Nevertheless, in the last 100 nucleotides of this region we did not find any *nodB*-like motif, and this is the zone where we have found two putative ORFs still open at the end of the reported sequence. An insertion at this point (pRp22) (7) does not cause any altered nodulation phenotype. Thus, we know that in this 400-nucleotide region there is neither a functional *nodB* gene nor any other essential nodulation genes. We have also found a *nod* box-like sequence within the *nodA* coding region (Fig. 4C). Although this sequence is not a perfect *nod* box, it precedes the *nodB*-like zones. The particular organization of the common *nod* genes in strain CE-3 could be the result of a complex rearrangement which occurred in a previous structure of these genes. Whatever its cause, it should have occurred before the diversification of the *R. leguminosarum* bv. phaseoli type I strains, as this organization seems to be a general phenomenon in this group of rhizobia, as suggested by our *nodABC* RFLP analysis (Table 3).

Both the *nodA* and ORF1-ORF2-*nodB*-*nodC* cistrons show *nod* boxes upstream of these genes (Fig. 4C). The distances from the *nod* boxes to the *nodA* and ORF1 initiation codons are 74 and 30 nucleotides, respectively. Both are shorter distances than those found in other *Rhizobium* and *Bradyrhizobium* species, in which the distance between the *nod* box and the initiation codon of the first ORF found at the common *nod* operon is between 150 and 200 bases (30, 46, 48). Both *nod* boxes are similar to the consensus sequence. Nevertheless, they are less conserved with respect to the consensus sequence than are the *nod* boxes found at the 5′ end of the *nodA* genes from diverse species.

Although the *nod* boxes in the two regions are different from one another, we could not find any differential expression of the *nodA* or *nodC* gene with the different inducers used. The two genes' general patterns of regulation seem to be similar when the fusions are harbored in multicopy plasmids in the wild-type strain CE-3 (Fig. 5). Expression of the *nodC-lacZ* fusion is in all cases lower than expression of the *nodA-lacZ* fusion. Thus, although the organization of the common *nod* genes is different in strain CE-3, the regulation seems not to be grossly affected. In consequence, it is probable that the same regulatory molecule(s) could be inducing or repressing the expression of these genes, giving as a final result a coordinated NodABC production, as found when these genes are organized as an operon. Nevertheless, we are still searching for a possible difference of regulation, exploiting the advantage of the natural separation of these genes in *R. leguminosarum* bv. phaseoli type I strains.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Academy of Sciences/National Research Council by means of grant BNF-MX-6877 from the Agency for International Development and by EEC grant CI1-0105-MEX.

We thank J. Miranda and O. Santana for technical assistance, L. Segovia for help in computer analysis, and A. Covarrubias for

critical review of the manuscript. M.V. was supported by a Consejo Nacional de Ciencia y Tecnología fellowship.

REFERENCES

1. **Bachem, C. W. B., E. Kondorosi, Z. Banfalvi, B. Horvath, A. Kondorosi, and J. Schell.** 1985. Identification and cloning of nodulation genes from the wide host range *Rhizobium* strain MPIK-3030. Mol. Gen. Genet. **199**:271–278.
2. **Banfalvi, Z., A. Nieuwkoop, M. Schell, L. Besl, and G. Stacey.** 1988. Regulation of *nod* gene expression in *Bradyrhizobium japonicum*. Mol. Gen. Genet. **214**:420–424.
3. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
4. **Brom, S., E. Martinez, G. Dávila, and R. Palacios.** 1988. Narrow- and broad-host-range plasmids of *Rhizobium* spp. strains that nodulate *Phaseolus vulgaris*. Appl. Environ. Microbiol. **54**:1280–1283.
5. **Casadaban, M. J.** 1975. Fusion of the *Escherichia coli lac* genes to the *ara* promoter: a general technique using bacteriophage Mu-1 insertions. Proc. Natl. Acad. Sci. USA **72**:809–813.
6. **Casadaban, M. J.** 1976. Transposition and fusion of the *lac* genes to selected promoters in *E. coli* using bacteriophage lambda and Mu. J. Mol. Biol. **104**:541–555.
7. **Cevallos, M. A., M. Vázquez, A. Dávalos, G. Espín, J. Sepúlveda, and C. Quinto.** 1989. Characterization of *Rhizobium phaseoli* Sym plasmid regions involved in nodule morphogenesis and host-range specificity. Mol. Microbiol. **3**:879–889.
8. **Davis, E. O., and A. W. B. Johnston.** 1990. Analysis of three *nodD* genes in *Rhizobium leguminosarum* biovar *phaseoli*; *nodD*1 is preceded by *nolE*, a gene whose product is secreted from the cytoplasm. Mol. Microbiol. **4**:921–932.
9. **Davis, E. O., and A. W. B. Johnston.** 1990. Regulatory functions of the three *nodD* genes of *Rhizobium leguminosarum* biovar *phaseoli*. Mol. Microbiol. **4**:933–941.
10. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
11. **Faucher, C., S. Camut, J. Denarié, and G. Truchet.** 1989. The *nodH* and *nodQ* host range genes of *Rhizobium meliloti* behave as avirulence genes in *R. leguminosarum* bv. *viciae* and determine changes in the production of plant-specific extracellular signals. Mol. Plant Microbe Interact. **2**:291–300.
12. **Figurski, D. H., and D. R. Helinski.** 1979. Replication of an origin containing derivative of plasmid RK2 dependent on a plasmid function provided *in trans*. Proc. Natl. Acad. Sci. USA **76**:1648–1652.
13. **Fisher, R. F., T. T. Egelhoff, J. T. Mulligan, and S. R. Long.** 1988. Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. Genes Dev. **2**:282–293.
14. **Fisher, R. F., and S. R. Long.** 1989. DNA footprint analysis of the transcriptional activator proteins NodD1 and NodD3 on inducible *nod* gene promoters. J. Bacteriol. **171**:5492–5502.
15. **Goethals, K., M. Gao, K. Tomekpe, M. Van Montagu, and M. Holsters.** 1989. Common *nodABC* genes in *Nod* locus 1 of *Azorhizobium caulinodans*: nucleotide sequence and plant-inducible expression. Mol. Gen. Genet. **219**:289–298.
16. **Gottfert, M., B. Horvath, E. Kondorosi, R. Putnoky, F. Rodríguez-Quiñones, and A. Kondorosi.** 1986. At least two *nodD* genes are necessary for efficient nodulation of alfalfa by *Rhizobium meliloti*. J. Mol. Biol. **191**:411–426.
17. **Hong, G. F., J. E. Burn, and A. W. B. Johnston.** 1987. Evidence that DNA involved in the expression of nodulation (*nod*) genes in *Rhizobium* binds to the product of the regulatory gene *nodD*. Nucleic Acids Res. **15**:9677–9690.
18. **Hooykaas, P. J. J., P. M. Klapwijk, M. P. Nuti, R. A. Schilperoort, and A. Rorsch.** 1977. Transfer of the *Agrobacterium tumefaciens* Ti plasmid to avirulent agrobacteria and to *Rhizobium ex planta*. J. Gen. Microbiol. **98**:477–484.
19. **Jacobs, T. W., T. T. Egelhoff, and S. R. Long.** 1985. Physical and genetic map of a *Rhizobium meliloti* nodulation region and nucleotide sequence of *nodC*. J. Bacteriol. **162**:469–476.
20. **John, M., J. Schmidt, U. Wieneke, H. D. Krüssmann, and J. Schell.** 1988. Transmembrane orientation and receptor-like structure of the *Rhizobium meliloti* common nodulation protein NodC. EMBO J. **7**:583–588.
21. **Johnson, D., L. E. Roth, and G. Stacey.** 1989. Immunogold localization of the NodC and NodA proteins of *Rhizobium meliloti*. J. Bacteriol. **171**:4583–4588.
22. **Kondorosi, E., J. Gyuris, J. Schmidt, M. John, E. Duda, B. Hoffman, J. Schell, and A. Kondorosi.** 1989. Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. EMBO J. **8**:1331–1340.
23. **Kosslak, R. M., R. Bookland, J. Barkei, H. E. Paaren, and E. R. Appelbaum.** 1987. Induction of *Bradyrhizobium japonicum* common *nod* genes by isoflavones isolated from *Glycine max*. Proc. Natl. Acad. Sci. USA **84**:7428–7432.
24. **Lerouge, P., P. Roche, C. Faucher, F. Maillet, G. Truchet, J. C. Promé, and J. Denarié.** 1990. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature (London) **344**:781–784.
25. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Martínez, E., M. A. Pardo, R. Palacios, and M. A. Cevallos.** 1985. Reiteration of nitrogen fixation gene sequences and specificity of *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. J. Gen. Microbiol. **131**:1779–1786.
28. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 433. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Mulligan, J. T., and S. R. Long.** 1985. Induction of *Rhizobium meliloti nod* expression by plant exudate requires *nodD*. Proc. Natl. Acad. Sci. USA **82**:6609–6613.
30. **Nieuwkoop, A. J., Z. Banfalvi, N. Deshmane, D. Gerhold, M. G. Schell, K. M. Sirotkin, and G. Stacey.** 1987. A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. J. Bacteriol. **169**:2631–2638.
31. **Noel, K. D., A. Sánchez, L. Fernández, J. Leemans, and M. A. Cevallos.** 1984. *Rhizobium phaseoli* symbiotic mutants with transposon Tn*5*. J. Bacteriol. **158**:148–155.
32. **Palacios, R., C. Quinto, H. de la Vega, M. Flores, L. Fernández, M. Hernández, T. Ballado, and G. Soberón.** 1983. General organization of nitrogen fixation genes in *Rhizobium phaseoli*, p. 164–168. *In* A. Pühler (ed.), Molecular genetics of bacteria-plant interactions. Springer Verlag KG, Berlin.
33. **Piñero, D., E. Martínez, and R. K. Selander.** 1988. Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. Appl. Environ. Microbiol. **54**:2825–2832.
34. **Ramakrishnan, N., R. K. Prakash, S. Shantharam, N. M. Duteau, and A. G. Atherly.** 1986. Molecular cloning and expression of *Rhizobium fredii* USDA 193 nodulation genes: extension of host range for nodulation. J. Bacteriol. **168**:1087–1095.
35. **Ratet, P., J. Schell, and F. J. de Bruijn.** 1988. Mini-Mu*lac* transposons with broad-host-range origins of conjugal transfer and replication designed for gene regulation studies in Rhizobiaceae. Gene **63**:41–52.
36. **Rolfe, B. G.** 1988. Flavones and isoflavones as inducing substances of legume nodulation. Biofactors **1**:3–10.
37. **Rolfe, B. G., and P. M. Gresshoff.** 1988. Genetic analysis of legume nodule initiation. Annu. Rev. Plant Physiol. Plant Mol. Biol. **39**:297–319.
38. **Rossen, L., A. W. B. Johnston, and J. A. Downie.** 1984. DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAB* and *C* required for root hair curling. Nucleic Acids Res. **12**:9497–9508.
39. **Rossen, L., C. A. Shearman, A. W. B. Johnston, and J. A. Downie.** 1985. The *nodD* gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nodABC* genes. EMBO J. **4**:3369–3373.
40. **Rostas, K., E. Kondorosi, B. Horvath, A. Simoncsits, and A.**

Kondorosi. 1986. Conservation of extended promoter regions of nodulation genes in *Rhizobium*. Proc. Natl. Acad. Sci. USA **83**:1757–1761.

41. **Sánchez, F., C. Quinto, M. Vázquez, H. Spaink, C. A. Wijffelman, M. A. Cevallos, A. de las Peñas, F. Campos, J. Padilla, and M. Lara.** 1988. The symbiotic association of *Phaseolus vulgaris* and *Rhizobium leguminosarum* bv. *phaseoli*, p. 370–375. *In* R. Palacios and D. P. S. Verma (ed.), Molecular genetics of plant-microbe interactions. Proceedings of the 4th International Symposium on Molecular Genetics of Plant-Microbe Interactions. American Phytopathological Society, St. Paul, Minn.
42. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
43. **Schlaman, H. R. M., H. P. Spaink, R. J. H. Okker, and B. J. J. Lugtenberg.** 1989. Subcellular localization of the *nodD* gene product in *Rhizobium leguminosarum*. J. Bacteriol. **171**:4686–4693.
44. **Schmidt, J. R., R. Wingender, M. John, U. Wieneke, and J. Schell.** 1988. *Rhizobium meliloti nod*A and *nod*B genes are involved in generating compounds that stimulate mitosis of plant cells. Proc. Natl. Acad. Sci. USA **85**:8578–8582.
45. **Schofield, P. R., and J. M. Watson.** 1986. DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. Nucleic Acids Res. **14**:2891–2903.
46. **Scott, K. F.** 1986. Conserved nodulation genes from the nonlegume symbiont *Bradyrhizobium* sp. (Parasponia). Nucleic Acids Res. **14**:2905–2919.
47. **Simon, R., U. Priefer, and A. Pühler.** 1983. A broad host range mobilization system for *in vivo* genetic engineering transposon mutagenesis in gram-negative bacteria. Bio/Technology **1**:784–791.
48. **Spaink, H. P., R. J. H. Okker, C. A. Wijffelman, E. Pees, and B. J. J. Lugtenberg.** 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1Jl. Plant Mol. Biol. **9**:29–37.
49. **Stormo, G. D., T. D. Schneider, and L. M. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **9**:2971–2996.
50. **Török, I., E. Kondorosi, T. Stepkowski, J. Posfai, and A. Kondorosi.** 1984. Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nucleic Acids Res. **12**:9509–9524.
51. **van Brussel, A. A. N., S. A. J. Zaat, H. C. J. C. Cremers, C. A. Wijffelman, E. Pees, T. Tak, and B. J. J. Lugtenberg.** 1986. Role of plant root exudate and Sym plasmid-localized nodulation genes in the synthesis by *Rhizobium leguminosarum* of Tsr factor, which causes thick and short roots on common vetch. J. Bacteriol. **165**:517–522.
52. **Vargas, C., L. J. Martínez, M. Megías, and C. Quinto.** 1990. Identification and cloning of nodulation genes and host specificity determinants of the broad host-range *Rhizobium leguminosarum* biovar *phaseoli* strain CIAT899. Mol. Microbiol. **4**:1899–1910.
53. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC vectors. Gene **33**:103–119.
54. **Zaat, S. A. J., C. A. Wijffelman, H. P. Spaink, A. A. N. Van Brussel, R. J. H. Okker, and B. J. J. Lugtenberg.** 1987. Induction of the *nodA* promoter of *Rhizobium leguminosarum* Sym plasmid PRL1JI by plant flavanones and flavones. J. Bacteriol. **169**:198–204.