MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 3, No. 5, pp. 308-316, 1990

Identification of *nodS* and *nodU*, Two Inducible Genes Inserted Between the *Bradyrhizobium japonicum nodYABC* and *nodIJ* Genes

Michael Göttfert, Susanne Hitz, and Hauke Hennecke

Mikrobiologisches Institut, Eidgenössische Technische Hochschule, CH-8092 Zürich, Switzerland.
Received 30 January 1990. Accepted 11 May 1990.

The so-called common nodulation (*nod*) gene cluster of *Bradyrhizobium japonicum* is characterized by a unique composition of genes that are arranged in the following order: *nodY*, *nodA*, *nodB*, *nodC*, *nodS*, *nodU*, *nodI*, *nodJ*. As reported here, the identification of the two new genes *nodS* and *nodU* resulted from the DNA sequencing of a 4.5-kilobase *nodC*-downstream region covering *nodS*, *nodU*, *nodI*, and *nodJ*. The predicted NodS, NodU, NodI, and NodJ proteins had the following respective amino acid (aa) lengths and molecular weights (M_r): 209 aa, M_r 23,405; 569 aa, M_r 62,068; 306 aa, M_r 34,127; and 262 aa, M_r 28,194. The 3′ end of *nodC* overlapped the 5′ end of *nodS* by 71 nucleotides. Using translational fusions of *lacZ* to *nodC*, *nodS*, and *nodU*, the expression of these genes was shown to be inducible by the isoflavone daidzein and depended on transcription from a DNA region farther upstream. These data and the adjacent location of all genes suggested the existence of a *nodYABCSUIJ* operon. The *nodI* and *nodJ* gene products shared about 70% sequence similarity with the corresponding *Rhizobium leguminosarum* bv. *viciae* proteins; NodI belongs to the family of ATP-binding proteins that are constituents of bacterial binding protein-dependent transport systems. By interspecies hybridization, DNA regions homologous to *nodSU* were detected in other strains of *Bradyrhizobium*. Likewise, *nodS*- and *nodU*-like genes were identified in *Rhizobium* sp. strain NGR234 (A. Lewin, E. Cervantes, W. Chee-Hoong, and W. J. Broughton, Molecular Plant-Microbe Interactions 3:317-326, 1990) in which *nodS* confers host specificity for *Leucaena leucocephala*. We constructed *B. japonicum nodS* and *nodU* mutants that, however, had no altered nodulation behavior on soybean, cowpea, mung bean, and Siratro.

Additional keywords: gene regulation, operon structure.

Bacteria of the genera *Rhizobium* and *Bradyrhizobium* are able to induce nitrogen-fixing nodules on the roots of their legume host plants. The genes essential for nodule formation (*nod*) can be grouped into common *nod* and host-specificity (*hsn*) genes. This classification is based on the ability of the common *nod* genes to be functionally interchangeable between different species, whereas the *hsn* genes do not cross-complement each other.

The majority of the nodulation genes share a similar mode of transcriptional regulation. The product of the regulator gene *nodD* binds to the *nod*-box, a conserved promoter region upstream of nodulation genes, and acts as a transcriptional activator in the presence of plant-derived flavonoid compounds (there are variations of this regulatory pathway in cases where a strain harbors several copies of *nodD*) (see Long 1989 for a recent review). The highly conserved common nodulation genes *nodA*, *nodB*, and *nodC* are absolutely essential for nodule formation. These genes are located next to each other in the fast-growing *Rhizobium* species *R. leguminosarum* bv. *viciae* (Rossen *et al.* 1984), the endosymbiont of *Vicia sativa* L., *R. l.* bv. *trifolii* (Schofield and Watson 1986), the endosymbiont of *Trifolium* species, *R. meliloti* (Török *et al.* 1984), the endosymbiont of *Medicago sativa* L., and in the slow-growing *Bradyrhizobium* species *Bradyrhizobium* sp. *Parasponia* (Scott 1986) and *B. japonicum* (Lamb and Hennecke 1986; Nieuwkoop *et al.* 1987).

The presence of the common nodulation genes *nodI* and *nodJ* downstream of *nodABC* was first described for *R. l.* bv. *viciae* (Evans and Downie 1986). Apparently, they are also present in *R. meliloti* and *R. l.* bv. *trifolii* (reviewed by Long 1989). Mutations in *nodIJ* of *R. l.* bv. *viciae* resulted in only a slight nodulation delay. Strain TOM of *R. l.* bv. *viciae* carries an additional nodulation gene *nodX* downstream of *nodJ* and within the same operon as *nodABCIJ* (Davis *et al.* 1988). This gene enables strain TOM to nodulate the primitive pea cultivar Afghanistan, which is not nodulated by European strains of *R. l.* bv. *viciae* lacking *nodX* (Davis *et al.* 1988). Thus, *nodX* is a determinant of host specificity cotranscribed with common *nod* genes.

The slow-growing species *Bradyrhizobium* sp. *Parasponia* carries an additional gene, termed *nodK*, upstream of *nodA* that is cotranscribed with it (Scott 1986). In the slow-growing strain 110 of *B. japonicum*, a gene (*nodY*) exists at a similar position (Banfalvi *et al.* 1988). However, a mutational analysis of *nodY* has shown that this gene is not essential for the nodulation of soybean (Göttfert *et al.* 1989). The *B. japonicum nodABC* genes as well as the *nodI* and *nodJ* genes have been identified previously by interspecies hybridization and partial DNA sequence determination (Lamb and Hennecke 1986; Nieuwkoop *et al.* 1987; Göttfert *et al.* 1989). While *nodIJ* of the above-mentioned fast-growing strains of *Rhizobium*

Address reprint requests to H. Hennecke: Mikrobiologisches Institut, Eidgenössische Technische Hochschule, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland.

Nucleotide and/or amino acid sequence data is to be submitted to GenBank as accession number JO3685.

© 1990 The American Phytopathological Society

are located immediately downstream of *nodABC*, there is a gap of more than 2 kilobases (kb) between *nodC* and *nodI* in *B. japonicum* (Lamb and Hennecke 1986; Nieuwkoop *et al.* 1987). This led to speculation that the region between *nodC* and *nodI* of *B. japonicum* might harbor additional, as yet unidentified nodulation genes. Therefore, the purpose of this study was to analyze in greater detail the region downstream of the *B. japonicum nodC* gene.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages. Bacterial strains used in this study are listed in Table 1. Plasmids are described in the text and in Figure 1. Phages M13mp18 and M13mp19 (Norrander *et al.* 1983) served as cloning vectors for subsequent sequencing of cloned *nod* DNA fragments.

Table 1. Bacterial strains used in this study

Designation	Relevant characteristics[a]	Origin or reference
Escherichia coli		
RR28	*hsdR hsdM recA pheS12* Sm^r	Hennecke *et al.* 1982
S17-1	*hsdR*, RP4-2*kan*::Tn*7 tet*::Mu (chromosomally located)	Simon *et al.* 1983
JM101	Δ(*lac-proAB*) F′ (*traD36 proAB*$^+$ *lacI*^{q}-ΔM15)	Messing 1983
Bradyrhizobium japonicum		
110*spc*4	Sp^r, referred to as wild type	Regensburger and Hennecke 1983
Δ440[b]	*nodS* deletion/replacement mutation, Sm^r	This study
434[b]	*nodU* insertion mutant, Km^r	This study
61A24		Nitragin[c]
61A76		NifTal[d]
USDA 123		USDA[e]
Bradyrhizobium species		
32H1 (*Crotalaria*)		Nitragin
ANU289 (*Parasponia*)		P. Gresshoff[f]
ATCC 10319 (*Lupinus*)		ATCC[g]
Rhizobium species		
NGR234		W. Broughton[h]
NGR234Δ1		W. Broughton
R. leguminosarum bv. *trifolii* ATCC 14480		ATCC
R. l. bv. *phaseoli* 8002		Lamb *et al.* 1982
R. meliloti AK631		A. Kondorosi[i]

[a] Sm, streptomycin; Sp, spectinomycin; Km, kanamycin; and ^{r}, resistant.
[b] The numbers of the mutants correspond to the numbers of the plasmid clones used for their construction (compare to Fig. 1).
[c] Nitragin Co., Milwaukee, WI.
[d] *Rhizobium* Germplasm Resource Center (NifTal), Paia, HI.
[e] United States Department of Agriculture, Beltsville, MD.
[f] P. Gresshoff, Plant Molecular Genetics (OHLD), Institute of Agriculture, University of Tennessee, Knoxville.
[g] American Type Culture Collection, Rockville, MD.
[h] W. Broughton, Laboratoire de Biologie Moléculaire des Plantes Supérieures, Université de Genève, Switzerland.
[i] A. Kondorosi, formerly of the Institute of Genetics, Biological Research Center, Hungarian Academy of Science, Szeged, Hungary.

Media and bacterial growth conditions. *Escherichia coli* was grown in Luria-Bertani medium (Miller 1972) at 37° C. *B. japonicum* was grown in peptone-salts-yeast extract (PSY) medium (Regensburger and Hennecke 1983) at 30° C. For plasmid selection antibiotics were used as described by Göttfert *et al.* (1989).

Recombinant DNA techniques. These were performed following established protocols as compiled by Maniatis *et al.* (1982).

DNA hybridizations. For filter-bound DNA, nitrocellulose BA85 (Schleicher and Schuell, Dassel, Federal Republic of Germany) was used. The radioactive probes were obtained as described by Feinberg and Vogelstein (1984). Hybridizations were done according to Maniatis *et al.* (1982). For interspecies hybridizations a temperature of 55° C was used.

DNA sequence analysis. The DNA sequence was established using the dideoxynucleotide chain termination method (Sanger *et al.* 1977) with the help of a DNA sequencer model 370A of Applied Biosystems (Foster City, CA). For computer-aided DNA and protein sequence analysis, the UWGCG (Genetics Computer Group of the University of Wisconsin, Madison, WI) software package

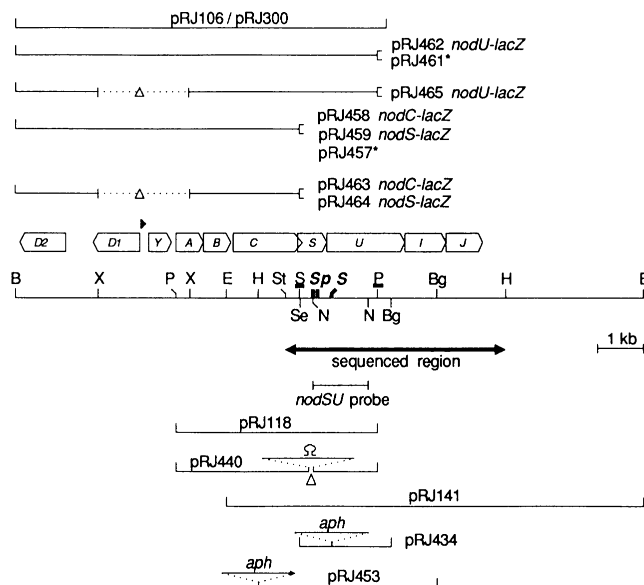

Fig. 1. Analysis of the *nodSUIJ* region downstream of the *Bradyrhizobium japonicum nodYABC* genes. The horizontal open arrows indicate the extent and direction of the *nod* genes as deduced from sequence analysis. The filled arrowhead between *nodD1* and *nodY* marks the location and orientation of the *nod*-box. The bold line below the restriction map marks the sequenced DNA fragment. For clarity only the restriction sites relevant to this work are included. Restriction sites used for mutagenesis and for constructing *lacZ* fusions are written in bold italics or underlined, respectively. Plasmid constructs and translational *nod′-′lacZ* fusions are shown above the gene symbols with Δ indicating the deleted *XhoI* fragment. Fusion constructs marked with an asterisk are out-of-frame fusions. Subclones used for the construction of mutants are shown in the lower part (the constructions are described in detail in the text). Abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; N, *Nsi*I; P, *Pst*I; S, *Sal*I; Se, *Spe*I; Sp, *Sph*I; St, *Stu*I; X, *Xho*I; *aph*, aminoglycoside phosphotransferase gene; Ω, omega interposon; and Δ, deletion.

(Release 6.1) was used. The percent identity values were determined by using the Gap program. Data base searches were done by using the TFastA program in conjunction with the data bases of GenBank (Release 61.0) and EMBL (Release 19.0).

Construction of plasmids for mutagenesis and constitutive expression of *nodS* and *nodU*. To mutagenize *nodS*, a 4.4-kb *Pst*I fragment carrying this region was first cloned into vector pHE3 (Hennecke *et al.* 1982), generating plasmid pRJ118 (Fig. 1). Then pRJ118 was cut with *Sph*I, which deletes a 93-base pair (bp) fragment between nucleotides 628 and 721 (compare Figs. 1 and 2), and the sticky ends were removed by T4 DNA polymerase treatment. The Ω interposon (Prentki and Krisch 1984) was inserted between these sites, and the enlarged *Pst*I fragment was cloned into the *Pst*I site of pSUP202, thus generating pRJ440 (Fig. 1). To mutagenize *nodU*, the 1.9-kb *Spe*I-*Bgl*II fragment (Fig. 1) isolated from pRJ141 (Göttfert *et al.* 1989) was cloned into the *Eco*RI site of pSUP202 (using *Eco*RI linkers). The *nodU* gene was disrupted by insertion of the kanamycin resistance cassette (derived from pUC4KIXX; Pharmacia, Uppsala, Sweden) into the *Sal*I site, thus generating pRJ434. Mobilization of pRJ440 and pRJ434 into *B. japonicum* and selection for marker exchange events were done as described previously (Göttfert *et al.* 1989) except that, in the case of pRJ440 (which carries an intact chloramphenicol resistance determinant), the *E. coli* donor was counter-selected by the addition of carbenicillin (40 μg/ml). The mutations in *nodS* and *nodU* were verified by Southern blot hybridizations.

To constitutively express *nodS* and *nodU* genes, pRJ141 was cut with *Hin*dIII and partially with *Bgl*II. The 4.2-kb *Hin*dIII-*Bgl*II fragment carrying the complete *nodS* and *nodU* genes was cloned into pPP375 (a gift of Peter Putnoky, Biological Research Center, Szeged, Hungary), a derivative of pRK290 carrying a multiple cloning site within the *Eco*RI restriction site. Finally, the kanamycin resistance cassette (derived from pUC4KIXX) was cloned into the *Hin*dIII site such that the resistance promoter led to the expression of the *nodSU* genes. This plasmid was named pRJ453 (Fig. 1).

Construction of translational *nod'-'lacZ* fusions. The 8-kb *Bam*HI insert of pRJ106 (Fig. 1; Lamb and Hennecke 1986) carrying most of the common *nod* region of *B. japonicum* was cloned into the *Bgl*II site of pRK290 (Ditta *et al.* 1980), resulting in plasmid pRJ300 (Fig. 1). Translational fusions of *nodC* and *nodS* to *lacZ* were generated by ligating the 5.1-kb *Sal*I-*Stu*I fragments (carrying the *E. coli lacZ* gene) of pNM480, pNM481, and pNM482 (Minton 1984) to the *Sal*I site in the *nodC*/*nodS* overlapping region of plasmid pRJ300, which resulted in plasmids pRJ457 (out-of-frame fusion), pRJ458 (in-frame *nodC'-'lacZ* fusion), and pRJ459 (in-frame *nodS'-'lacZ* fusion) (see Fig. 1). The translational *nodU'-'lacZ* fusions were constructed by fusing the 5.1-kb *Pst*I-*Stu*I fragments (carrying the *lacZ* gene) of pNM481 and pNM482 to the *Pst*I site within *nodU* of pRJ300, resulting in plasmids pRJ461 (out-of-frame fusion) and pRJ462 (in-frame fusion) (Fig. 1). The fusion sites of *B. japonicum* DNA and the *lacZ* gene were verified by DNA sequence analysis. For the purpose of removing the putative promoter region containing the *nod*-box, the same 2-kb *Xho*I fragment of each in-frame *nod'-'lacZ* fusion construct was deleted, thus generating plasmids pRJ463 (*nodC'-'lacZ* fusion), pRJ464 (*nodS'-'lacZ* fusion), and pRJ465 (*nodU'-'lacZ* fusion) (Fig. 1). All *nod'-'lacZ* fusion plasmids were transferred into *B. japonicum* by triparental mating as described by Ditta *et al.* (1980).

β-Galactosidase assays. Ten ml of PSY medium plus tetracycline (60 μg/ml) was inoculated with a single colony of *B. japonicum* harboring the *lacZ* fusion plasmid. At a cell density of $A_{600} = 0.8–1.0$, 160-μl samples of the culture were transferred to 10 ml of fresh PSY plus tetracycline medium. The cultures were grown for 24 hr, and inducer (daidzein) was added to a final concentration of 1 μM. After additional incubation for 20 to 24 hr, β-galactosidase activities were measured as described by Miller (1972) and modified by Fischer *et al.* (1986).

Plant infection tests. Soybean seeds (*Glycine max* (L.) Merr. cv. Williams) were provided by the Jacques Seed Co., Prescott, WI. *G. soja* Siebold & Zucc. (PI468397) seeds were provided by the U.S. Department of Agriculture, Beltsville, MD. Seeds from cowpea (*Vigna unguiculata* (L.) Walp. cv. Red Caloona), mung bean (*V. radiata* (L.) R. Wilcek), *Macroptilium atropurpureum* Urb. cv. Siratro, and *Leucaena leucocephala* (Lam.) de Wit cv. Cunningham were kindly provided by W. D. Broughton (Université de Genève, Switzerland). The seeds (except those of *L. leucocephala* that were surface-sterilized and germinated as described by Lewin *et al.* [1990]) were surface-sterilized by soaking them for 5 min in sodium hypochlorite (0.7%) and subsequently washing them four times with sterile water. Seeds were germinated for 2 days on water agar and in the dark. The seedlings were planted in 200-ml sterile glass jars containing vermiculite and 30 ml of nitrogen-free Jensen medium (Vincent 1970), and inoculated by adding 1 ml of bacterial suspension (approximately 10^8 bacteria). Plants were grown in a climatized chamber (Heraeus-Vötsch Ecophyt type VEPHL 5/1350) under the following conditions: 16-hr light and 8-hr dark period, 26° C and 80% relative humidity in the light, and 22° C and 80% relative humidity in the dark. A total of 16 plants of each species were inoculated with each bacterial strain, and eight plants were scored for nodules at days 12 and 21 (soybean) or days 14 and 21 (all other host plant species). This procedure allowed us to detect a nodulation delay of 2 or more days as well as a reduction in nodule number.

RESULTS

DNA sequence of the *nodS*, *nodU*, *nodI*, and *nodJ* genes. To analyze the region downstream of the *B. japonicum nodC* gene, we determined the nucleotide sequence of a 4.8-kb *Stu*I-*Hin*dIII fragment (Figs. 1 and 2). Nucleotides 1 to 317 represent the 3′ end of *nodC*. Downstream of *nodC* we identified four open reading frames (ORFs) that were subsequently named *nodS* (nucleotides 247 to 873), *nodU* (nucleotides 890 to 2596), *nodI* (nucleotides 2601 to 3518), and *nodJ* (nucleotides 3525 to 4310). The predicted

molecular weights of the deduced gene products are 23,405 (NodS), 62,068 (NodU), 34,127 (NodI), and 28,194 (NodJ). The positions of the *nodS* start codon at nucleotide 247 and the *nodC* stop codon at nucleotide 318 imply that the two ORFs of *nodC* and *nodS* overlap by as many as 71 bp, a highly unusual situation in bacteria. To exclude the possibility that this overlap was artificially created by a sequencing error or an accidental mutation picked up during the cloning procedures, we analyzed the *nodC*/*nodS* overlap from a second cosmid clone derived from an independent cosmid library and from a *Stu*I-*Spe*I fragment isolated from total DNA and obtained the identical sequence as before. The *nodU* ORF is preceded by a purine-rich region that (on the mRNA level) could be involved

```
   1  AGGCCTATGCGCTTTGTACATTGAGCAATAGCGATTGGCTATCTCGGAAAGTCACCGATATGCCGACGGAAGAGGGGAAACAGCCTGTCA
        A  Y  A  L  C  T  L  S  N  S  D  W  L  S  R  K  V  T  D  M  P  T  E  E  G  K  Q  P  V  I

  91  TCCTGCACCCGAATGCCGGACGAAGTCCTGCTGGTGTAGGGGGCGCCTGCTCCTATTCGTAAGGCGGCGTTATCGCAGCCTCCATCGAG
        L  H  P  N  A  G  R  S  P  A  G  V  G  G  R  L  L  L  F  V  R  R  R  Y  R  S  L  H  R  A

 181  CCTGGCGGCGACGGAGAGTGTTTCCGGTCGCGATCGTTCGACTGTCTACAAATAAGTGGTCGGCTGATGACTCAGGACGAAAACCATCAG
        W  R  R  R  R  V  F  P  V  A  I  V  R  L  S  T  N  K  W  S  A  D  D  S  G  R  K  P  S  V
                                                                         M  T  Q  D  E  N  H  Q

 271  TTATTAGAGCGAGAGTTGGCTGTCGACGACCCGTGGCGCCTCGACACTAGTGCGTTCGAGCAGCAGCGATACGCGCAAATGCTCCGGATG
        I  R  A  R  V  G  C  R  R  P  V  A  P  R  H  *
       L  L  E  R  E  L  A  V  D  D  P  W  R  L  D  T  S  A  F  E  Q  Q  R  Y  A  Q  M  L  R  M

 361  TCGCGCCGCGACGGAGATGCGGCGTCTGCGCTCGAAGTAGGATGTGCAGCCGGTGCATTCACGGAGATGCTGGCGCCGCTATGCGAGCGA
       S  R  R  D  G  D  A  A  S  A  L  E  V  G  C  A  A  G  A  F  T  E  M  L  A  P  L  C  E  R

 451  CTTACGGTCGTGGATGTCATGCCGCAGGCAATCGAGCGAACGCGACTGCGGACCGGGAAATGGTCACATATTAGTTGGGTGACCTGCGAC
       L  T  V  V  D  V  M  P  Q  A  I  E  R  T  R  L  R  T  G  K  W  S  H  I  S  W  V  T  C  D

 541  ATTCAGCGGTTCTCGACCACTGAGCAATTCGATTTGATTGTCGTGGCTGAGGTTCTTTACTACCTCAGGGACGTCGTCGAGATGCATGCG
       I  Q  R  F  S  T  T  E  Q  F  D  L  I  V  V  A  E  V  L  Y  Y  L  R  D  V  V  E  M  H  A

 631  GCTATCCGCAACCTGGTGAGTATGCTTGCGCCAGATGAGACTCTCATTTTCGGGTCTGCGCGTGATGAAATATGTCAACGCTGGGGGCAT
       A  I  R  N  L  V  S  M  L  A  P  D  E  T  L  I  F  G  S  A  R  D  E  I  C  Q  R  W  G  H

 721  GCTGCCGGTGCTGAGACGGTGATCGCTCTCTTCAATGAGAGCCTATCAGAGATCGAGCGCCGGCACTGCTGCACCGGGTCGGCGAGCGAA
       A  A  G  A  E  T  V  I  A  L  F  N  E  S  L  S  E  I  E  R  R  H  C  C  T  G  S  A  S  E

 811  GACTGCTTGATCGTCCGGTTTCTGAAGCCGGGCGGCGCGTCGGCACAATCAAACGCCGGTCATTAGGCCAGGAGGACGTATGCGCATCTG
       D  C  L  I  V  R  F  L  K  P  G  G  A  S  A  Q  S  N  A  G  H  *                 M  R  I  C

 901  CGGAATAAAGTTGACACATGACGGAGCAATTGCTGTCGTCGAGGACGGTCGGCGTCTTTTTTGCGTCGAGCAAGAGAAGCGCGGCAATGG
        G  I  K  L  T  H  D  G  A  I  A  V  V  E  D  G  R  R  L  F  C  V  E  Q  E  K  R  G  N  G

 991  TCCTCGCTATCAGTCCGTCGACAATCTCGATGCAGTCGTCTTCGCCTTGGCGGAGCATGGCCTGAATCCGCGCGACATCGATCAGTTCGT
        P  R  Y  Q  S  V  D  N  L  D  A  V  V  F  A  L  A  E  H  G  L  N  P  R  D  I  D  Q  F  V

1081  GATTGATGGCTGGGACGGGGAGAATGAATCGCAGTTCCAGCTCCTAAGCGGAGCGGTGCCGGTCGCGCTAAAAGGCGCGCCATATGTCGA
        I  D  G  W  D  G  E  N  E  S  Q  F  Q  L  L  S  G  A  V  P  V  A  L  K  G  A  P  Y  V  E

1171  GCGTCATGCTGAGGGCCTCCTTGATTCCGTCGACGGCTATGGCCTCCTCCTCGGAGGCGAGGAGTTTCCATATAAGAGCTACCCGCATGT
        R  H  A  E  G  L  L  D  S  V  D  G  Y  G  L  L  L  G  G  E  E  F  P  Y  K  S  Y  P  H  V

1261  CACGGGCCATGTCGCCTCAGCATATAGCACCAGTCCTTTCGCCAGCGCGGGGAAACCCGCGTTGTGTCTGGTATGGGACGGCTGTATCTT
        T  G  H  V  A  S  A  Y  S  T  S  P  F  A  S  A  G  K  P  A  L  C  L  V  W  D  G  C  I  F

1351  TCCACGCCTTTACTATGTCGAGCCCCAGGGCGCGCGGCTCATCGGATCGCTGTTTCCGATGATTGGCCATGCCTATGCCGCCGCGGGCCT
        P  R  L  Y  Y  V  E  P  Q  G  A  R  L  I  G  S  L  F  P  M  I  G  H  A  Y  A  A  A  G  L

1441  TCACTTCGGCCCATACCGGCAGCCGAACCGCTCCAGTTGGGATTTGGGCATTGCTGGCAAGCTGATGGCCTACATCGAGCTTGGTTCCGT
        H  F  G  P  Y  R  Q  P  N  R  S  S  W  D  L  G  I  A  G  K  L  M  A  Y  I  E  L  G  S  V

1531  GGACGAAAGCATCGTGGAAGTGTTCCAGGGGCTTTATGAAACGCGCTCAGCGGCTGACACGGAGCAGGCTCGTCGCTACCGCGAAAACAT
        D  E  S  I  V  E  V  F  Q  G  L  Y  E  T  R  S  A  A  D  T  E  Q  A  R  R  Y  R  E  N  I

1621  CAACAATGCGGAAGCGTCTCTCGCAGTTATTCACGACTTCTTCGAGTCAAGCGCATTGCGCCTGAAGGCCAAGCGCGCGGAGGACGTCCT
        N  N  A  E  A  S  L  A  V  I  H  D  F  F  E  S  S  A  L  R  L  K  A  K  R  A  E  D  V  L

1711  CGCGTCGTTTCATGTGTTTCTGGAACGTCTTCTCGTTAAGGAGATCGCGATGGTTCTGCTACGGCACTCGTCGCTTCCGGGAGCGCGAAA
        A  S  F  H  V  F  L  E  R  L  L  V  K  E  I  A  M  V  L  L  R  H  S  S  L  P  G  A  R  N

1801  TCTATGCATCGCCGGAGGTTGCGGTCTCAACATCAAATGGAACAGCGCGCTTCGCGCGACCGGATTATTCGATGATGTCTGGGTGCCGCC
        L  C  I  A  G  G  C  G  L  N  I  K  W  N  S  A  L  R  A  T  G  L  F  D  D  V  W  V  P  P

1891  GTTTCCAAATGACAGTGGCTCGGCAATCGGCGCGGCGTGCGGCGCTATGGCAGCGCAAGATGGCTTCGAGCCGTTGGAATGGTCAGTTTA
        F  P  N  D  S  G  S  A  I  G  A  A  C  G  A  M  A  A  Q  D  G  F  E  P  L  E  W  S  V  Y

1981  CAGTGGCCCCGCCCTGCAGGAGAGCGAGGTTCCGCCGGACTGGGAGGCGGCGCCGTGCAGTCTGCCTGAACTTGCCTCGATTCTTGCTGA
        S  G  P  A  L  Q  E  S  E  V  P  P  D  W  E  A  A  P  C  S  L  P  E  L  A  S  I  L  A  D

2071  CAACAAGCCCGTCATCTTTCTTTCCGGGTGTGCCGGGCTCGGGCCGCGGGCGTTGGGCGGTAGAAGCATTCTTGCGGCCCCAACGTCCCC
        N  K  P  V  I  F  L  S  G  C  A  G  L  G  P  R  A  L  G  G  R  S  I  L  A  A  P  T  S  P

2161  GGAGATGAAGGATCATCTCAACGACATCAAACGGCGCGAACACTTCCGGCCGGTGGTGCCGATCTGTCTGGAAGATCGTGCGCCGGAGAT
        E  M  K  D  H  L  N  D  I  K  R  R  E  H  F  R  P  V  V  P  I  C  L  E  D  R  A  P  E  I

2251  CTTCAGCCCGGGTACGCCAGATCCTTACATGCTATTCGATCACCAAACCCGGGCGAATTGGCGCGACAAGATCCCGGCAGTGGTACATCT
        F  S  P  G  T  P  D  P  Y  M  L  F  D  H  Q  T  R  A  N  W  R  D  K  I  P  A  V  V  H  L

2341  CGACGGCTCGGCGCGGCTGCAGACAATTTCCCGCAACTCTCCTCACAAAATTGCTGCGCTTCTCATCGAATTTGAGCAACTCACCGGCAT
        D  G  S  A  R  L  Q  T  I  S  R  N  S  P  H  K  I  A  A  L  L  I  E  F  E  Q  L  T  G  I
```

Fig. 2. Nucleotide sequence of the *nodSUIJ* region. The sequence shown begins with the 3′ end of *nodC* (nucleotides 1 to 317) and continues with *nodS* (nucleotides 247 to 873), *nodU* (nucleotides 890 to 2596), *nodI* (nucleotides 2601 to 3518), and *nodJ* (nucleotides 3525 to 4310) plus flanking DNA on the 3′ side of *nodJ*. The derived amino acid sequences of the gene products are shown in one-letter code. (**Fig. 2.** *continued on next page.*)

in ribosome binding. The start of the *nodI* ORF is characterized by the presence of three ATG codons within a stretch of 18 nucleotides. We do not know which of them serves as the translational start codon for *nodI* and have arbitrarily chosen the first ATG at position 2601. The putative start codon of *nodJ* is preceded by a Shine-Dalgarno-like sequence (Shine and Dalgarno 1974), which overlaps with the 3′ end of *nodI*. No secondary structure potentially able to serve as a transcriptional stop signal was identified downstream of *nodJ*. Thus, we cannot exclude the possibility that a presumptive operon extends even farther beyond *nodJ*.

Interspecies hybridizations and amino acid sequence comparisons of the deduced *nodS* and *nodU* gene products. In previous interspecies DNA hybridization experiments (M. Göttfert, unpublished), we observed that the *nodSU* region of *B. japonicum* and the so-called HsnII locus of *Rhizobium* sp. strain NGR234 may share some homology. This rhizobial strain is able to nodulate a wide range of (mainly tropical) legume host plants and has been shown to carry at least three separate DNA regions conferring host specificity (Hsn; Lewin *et al.* 1987). Because of the positive, albeit weak, interspecies hybridization between the NGR234 HsnII locus and *B. japonicum nodSU* genes,

```
2431  TCCGCTGCTCTGCACCACGAGCGCCAACCTCCATGGACGGGGATTCTTTCCCGATGCTGCGGCGGCATGCCAATGGGGGCGCGTCGAGCA
        P  L  L  C  T  T  S  A  N  L  H  G  R  G  F  F  P  D  A  A  A  A  C  Q  W  G  R  V  E  H

2521  TGTGTGGTGCGAGGGCATGCTCTGGAGCAAAACGGTCATCAAGAAGTCATCGCCTACGGAACGACTTCTGTCAGCCTAAGATGAACATGT
        V  W  C  E  G  M  L  W  S  K  T  V  I  K  K  S  S  P  T  E  R  L  L  S  A  *     M  N  M  S

2611  CCAATATGGCGATCGACCTTGTCGGCGTGAGAAAGTCATTCGGGGACAAGGTTATCGTCAATGACCTGTCGTTCTCGGTTGCGCGCGGAG
        N  M  A  I  D  L  V  G  V  R  K  S  F  G  D  K  V  I  V  N  D  L  S  F  S  V  A  R  G  E

2701  AGTGCTTCGGGCTGCTTGGGCCAAATGGAGCTGGCAAAAGTACGATTGCACGCATGCTCCTCGGCATGATTTCCCCCGATCGAGGCAAGA
        C  F  G  L  L  G  P  N  G  A  G  K  S  T  I  A  R  M  L  L  G  M  I  S  P  D  R  G  K  I

2791  TTACGGTCCTCGATGAGCCTGTGCCTTCCCGCGCTCGTGCGGCGCGCGTCGCCGTAGGCGTGGTGCCGCAGTTCGATAACCTTGAGCCCG
        T  V  L  D  E  P  V  P  S  R  A  R  A  A  R  V  A  V  G  V  V  P  Q  F  D  N  L  E  P  E

2881  AGTTCACCGTGCGGGAGAACCTGCTTGTGTTTGGCCGCTATTTCGGCATGAGCGCTCGCACGATCGAAGCGGTTGTGCCCTCGTTGCTTG
        F  T  V  R  E  N  L  L  V  F  G  R  Y  F  G  M  S  A  R  T  I  E  A  V  V  P  S  L  L  E

2971  AGTTTGCGCGGCTTGAAAGCAAGGCGGACGTGCGCGTCTCCCTTTTGTCCGGTGGTATGAAGCGGCGGCTGACGCTGGCGCGCGCTCTGA
        F  A  R  L  E  S  K  A  D  V  R  V  S  L  L  S  G  G  M  K  R  R  L  T  L  A  R  A  L  I

3061  TCAATGATCCACATCTACTCGTGATGGACGAGCCGACGACGGGGCTCGATCCGCATGCTCGCCACCTGATCTGGGAACGCCTGCGGGCTC
        N  D  P  H  L  L  V  M  D  E  P  T  T  G  L  D  P  H  A  R  H  L  I  W  E  R  L  R  A  L

3151  TTCTTGCGCGCGGCAAGACGATCCTCTTGACCACTCACTTCATGGAAGAGGCCGAACGCCTTTGCGATCGGCTATGCGTGCTTGAGAGTG
        L  A  R  G  K  T  I  L  L  T  T  H  F  M  E  E  A  E  R  L  C  D  R  L  C  V  L  E  S  G

3241  GATGCAAAATCGCCGAAGGCAAGCCAGACGCCTTGATCGACGAGCATATCGGCTGCAACGTGATCGAGATCTATGGCGGTGATCTAGATC
        C  K  I  A  E  G  K  P  D  A  L  I  D  E  H  I  G  C  N  V  I  E  I  Y  G  G  D  L  D  Q

3331  AACTCCGGGAGCTGATCAGGCCGTATGCGCGGCATATCGAAGTGAGTGGAGAGACGCTTTTTTGTTACGCGCGATGTCCGGACGAAATCA
        L  R  E  L  I  R  P  Y  A  R  H  I  E  V  S  G  E  T  L  F  C  Y  A  R  C  P  D  E  I  S

3421  GCGTGCACCTGCGCGGGCGAACGGACCTTCGCGTTCTGCAGCGCCCCCCGAATCTCGAAGACGTGTTTTTGCGGTTGACCGGGCGCGAGA
        V  H  L  R  G  R  T  D  L  R  V  L  Q  R  P  P  N  L  E  D  V  F  L  R  L  T  G  R  E  M

3511  TGGAGAAATGAGCGATGGATGATGGTTATGCGTCGGTCATGCCGGCTAATGCGTACAATTGGACCGCGGTATGGCGTCGAAATTACTTGG
        E  K  *     M  D  D  G  Y  A  S  V  M  P  A  N  A  Y  N  W  T  A  V  W  R  R  N  Y  L  A

3601  CATGGAGGAAAGTCGCGCTTGCATCGCTTCTCGGCAACCTCGCAGATCCCATAACCAATCTGTTTGGCCTTGGCTTTGGGCTCGGACTCA
        W  R  K  V  A  L  A  S  L  L  G  N  L  A  D  P  I  T  N  L  F  G  L  G  F  G  L  G  L  I

3691  TTGTGGGACGGGTTGAGGGGACTTCGTACATTGCGTTTTTGGCGGCGGGTATGGTCGCGATTAGCGCTATGACATCCGCGACCTTTGAAA
        V  G  R  V  E  G  T  S  Y  I  A  F  L  A  A  G  M  V  A  I  S  A  M  T  S  A  T  F  E  T

3781  CTCTGTACGCAGCCTTTGCTCGGATGGATGTCAAGCGCACCTGGGAGGGAATTTTGTTCACACAGCTCACGCTTGGCGATATCGTTCTAG
        L  Y  A  A  F  A  R  M  D  V  K  R  T  W  E  G  I  L  F  T  Q  L  T  L  G  D  I  V  L  G

3871  GTGAGTTGGTGTGGGCGGCCAGTAAGTCCGTTCTAGCCGGGACAGCAATCGGGATTGTCGCTGCCACCCTGGGTTATGCATCCTGGACGT
        E  L  V  W  A  A  S  K  S  V  L  A  G  T  A  I  G  I  V  A  A  T  L  G  Y  A  S  W  T  S

3961  CGGTTCTTTGTGCGATACCGACAATCGCCCTTACGGGCCTTGTCTTCGCCAGCCTGGCAATGGTCGTCATATCTCTTGCGCCAACTTACG
        V  L  C  A  I  P  T  I  A  L  T  G  L  V  F  A  S  L  A  M  V  V  I  S  L  A  P  T  Y  D

4051  ATTACTTCGTGTTTTACCAGTCGCTCGTCCTTACGCCCATGGTGTTCTTGTGTGGCGCGGTCTTTCCGACGAGCCAAATGCCCGACTCAT
        Y  F  V  F  Y  Q  S  L  V  L  T  P  M  V  F  L  C  G  A  V  F  P  T  S  Q  M  P  D  S  F

4141  TTCAGCACTTTGCCGGCTTGTTGCCGCTGGCACATTCGGTCGACCTTATTCGCCCAGTGATGCTTGAGCGCGGCGCCGACAATGCCGCCC
        Q  H  F  A  G  L  L  P  L  A  H  S  V  D  L  I  R  P  V  M  L  E  R  G  A  D  N  A  A  L

4231  TACACGTAGGCGCGCTCTGCGTTTACGCGGTCTTGCCCTTTTTCGCGTCGATCGCGTTATTTCGGCGCCGCCTGCTGCGTTGACGTGAGT
        H  V  G  A  L  C  V  Y  A  V  L  P  F  F  A  S  I  A  L  F  R  R  R  L  L  R  *

4321  TCAGCGAAACGGAGAAGTCAAAGCGATGCAGTGTCATGATGCGTTCCGCGGTAGCTCGCGAGTGACCTGGTCTGTCAGAGCAGCTTGTGA

4411  AACGGCCTTGAGAAGCCGCGATCGAGCCAGCTCTACGGCCGCCTGTTCGGCCGATCGGACCTTGCAAAGTCGTGTTGTGCTGCCTGGCAC

4501  GCGCACCGGCTGCGCGTGACTGTACAGATTAGATCGGTGGGACGGCATGGCCTTCCATGGAGAACGGGATGTTGAGCATAGGACTGAGTG

4591  GCGGCCTTGACAGAATCTATGAAAGCTCTCCGGAGCTGCCGAATACATTTCTTCACGATGGCGCTGCGGTACTTGTGCAGGACGGCCGCG

4681  TAATTGCTGCTGTCGAAGAGGAGCGCCTCAACCGCGTCAAGCACTCCAACAAGTTCCCCAGCAATTCGATCCGATACTGCCTTTCTACCG

4771  CAGGAGTCGAGCTCGGCGACATCGACCGTATCGCGTTCTACGCGACTGAAGCTT   4824
```

Fig. 2 *continued from previous page.*

we were interested in comparing the DNA sequences of both regions. As discovered by Lewin *et al.* (1990), strain NGR234 has a *nodS* gene as well. Figure 3 shows an amino acid sequence comparison of the predicted NodS proteins of both species, which revealed that they shared 60% identical amino acid residues at corresponding positions. Neither the NodS protein nor the NodU protein of *B. japonicum* showed any significant similarities to other proteins present in the sequence data bank. In *Rhizobium* sp. strain NGR234 there is now evidence to suggest that this strain also carries a *nodU*-like gene immediately adjacent to the 3′ end of *nodS* (Lewin 1989; Lewin *et al.* 1990).

Having found *nodSU* homology between *B. japonicum* and *Rhizobium* sp. strain NGR234, we were prompted to look for similar sequences in other strains of *Bradyrhizobium* and *Rhizobium*. A 1.2-kb *Nsi*I fragment of the *nodSU* region (Fig. 1) was used as a radioactive probe and hybridized to total DNA of various strains. As can be seen in Figure 4, hybridization is generally strong with DNA of *Bradyrhizobium* strains (lanes 2–5, 8, 10, and 11) but weak (NGR234, lane 1) or absent with DNA of *Rhizobium* strains (lanes 6, 7, and 12).

Amino acid sequence comparisons of the deduced *nodI* and *nodJ* gene products. The amino acid sequence comparison of the predicted *nodI* and *nodJ* gene products from *B. japonicum* to the corresponding NodI and NodJ proteins of *R. l.* bv. *viciae* showed a high degree of homology (Fig. 5). NodI and NodJ proteins of both species shared 73 and 66% identical amino acid residues, respectively. A data base search revealed that NodI was a member of the large family of ATP-binding proteins, which confirmed the finding first made by Evans and Downie (1986). All members of this family (reviewed by Higgins *et al.* 1988) share amino acid sequence similarities of 20 to 30% within a domain of approximately 200 amino acids. The *nodJ* gene product is a very hydrophobic protein and, thus, is likely to be membrane-bound. NodJ showed no significant similarity to any other protein present in the data bank.

Regulation of *nodS* and *nodU* expression. The location of *nodS* and *nodU* immediately downstream of *nodC* suggests that these genes are located within the same operon

```
Bjap      1                  MTQDENHQLLERELAVDDPWRLDTSAFEQ
                              ||  |  || ||||  ||||||  ||||
NGR234    1  MCKSLCRSVHGVSEANLTQVNNYHLLHRELAAEDPWRLDANAFEQ

Bjap     30  QRYAQMLRMSRRDGDAASALEVGCAAGAFTEMLAPLCERLTVVDV
              |  |||| |   |    ||||||||||||| ||| | |||| ||
NGR234   46  ERHSQMLRLSLSQGPITNALEVGCAAGAFTEKLAPYCKRLTVIDV

Bjap     75  MPQAIERTRLRTGKWSHISWVTCDIQRFSTTEQFDLIVVAEVLYY
              | || ||| |    ||| |   |   |||   ||||||||||||
NGR234   91  VPRAIARTRQRMKESSHINWIVADVRQFSTQQLFDLIVVAEVLYY

Bjap    120  LRDVVEMHAAIRNLVSMLAPDETLIFGSARDEICQRWGHAAGAET
             | ||     |  ||||||||     |||| |  | |||| |||||
NGR234  136  LEDVAAIRTAVHNLVSMLAPSGHMVFGSAIDANCRRWGHVAGAET

Bjap    165  VIALFNESLSEIERRHCCTGSASEDCLIVRFLKPGGASAQSNAGH
             |||  || | | ||  |   |  ||||  || |
NGR234  181  VIAMLNETLIEVERLYCRGASVNEDCLLSRFQKSTT
```

Fig. 3. Amino acid sequence alignment of the NodS proteins of *Bradyrhizobium japonicum* (Bjap; compare with Fig. 2) and *Rhizobium* sp. strain NGR234 (Lewin *et al.* 1990). Identical amino acids are connected by vertical lines.

as *nodYABC*. This would imply that the expression of *nodS* and *nodU* is regulated by *nodD* in conjunction with the *nod*-box in front of *nodY* (Fig. 1) and that transcription is inducible by the isoflavone daidzein, which is known to be an inducer of *nod* gene expression in *B. japonicum* (Kosslak *et al.* 1987; Banfalvi *et al.* 1988; Göttfert *et al.* 1988). To test this hypothesis, we constructed translational *nod′-′lacZ* fusions (Fig. 1). For the *nodC′-′lacZ* and *nodS′-′lacZ* fusions, we chose the *Sal*I restriction site at nucleotide position 292 (Fig. 2). This site is particularly useful because fusing the *lacZ* gene in all three frames to this site creates either an in-frame *nodC′-′lacZ* fusion, an in-frame *nodS′-′lacZ* fusion, or an out-of-frame fusion. Moreover, this site is within that part of the 3′ end of *nodC* which has no equivalent in *nodC* of *R. meliloti* or *R. l.* bv. *viciae*. For the construction of an in-frame and an out-of-frame *nodU′-′lacZ* fusion, the *Pst*I site at position 1998 was chosen. As shown in Table 2, the expression of all in-frame *lacZ* fusions is inducible by daidzein. We also tested the dependence of *nod′-′lacZ* gene expression on the presence of upstream sequences. For this purpose a 2-kb *Xho*I fragment was deleted, removing the *nod*-box and flanking DNA regions (Fig. 1). The β-galactosidase activities obtained with all deletion derivatives had only

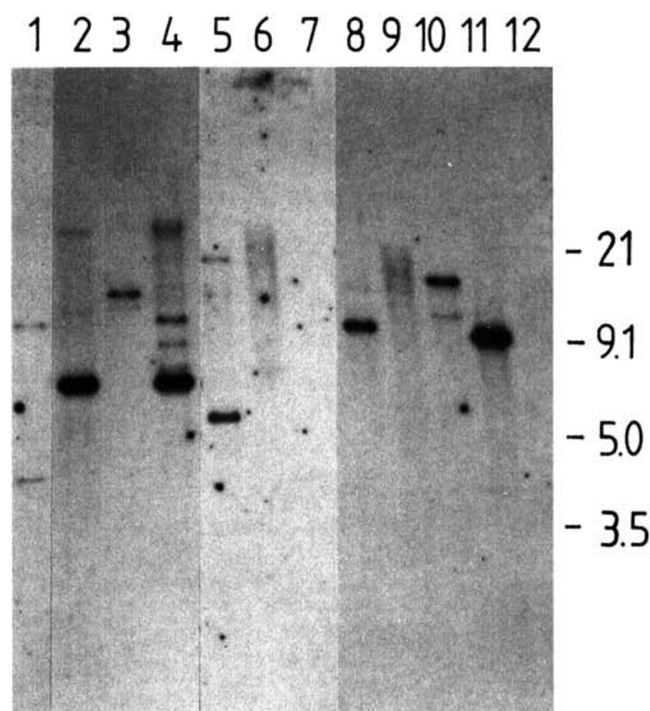

Fig. 4. *nodSU*-specific interspecies hybridization to DNA of different strains of *Bradyrhizobium* and *Rhizobium*. The autoradiograph was obtained after hybridization of *Eco*RI-cut, filter-bound total DNA of different strains to a 1.2-kilobase *nodSU* probe (compare with Fig. 1). Numbers in the right margin are size markers (in kilobases). Total genomic DNA was from the following sources: lane 1, *Rhizobium* sp. strain NGR234; lane 2, *B. japonicum* 61A24; lane 3, *B. japonicum* 61A76; lane 4, *B. japonicum* USDA 123; lane 5, *Bradyrhizobium* sp. strain ATCC 10319; lane 6, *R. leguminosarum* bv. *trifolii* ATCC 14480; lane 7, *R. l.* bv. *phaseoli* 8002; lane 8, *Bradyrhizobium* sp. strain 32H1; lane 9, *Escherichia coli* RR28; lane 10, *Bradyrhizobium* sp. strain ANU289; lane 11, *B. japonicum* 110*spc*4; and lane 12, *R. meliloti* AK631.

background values, and the expression was no longer inducible by the addition of daidzein (Table 2). This shows that the expression of *nodC*, *nodS*, and *nodU* is dependent on the same upstream sequence, most probably the *nod*-box.

Nodulation phenotypes of *nodS* and *nodU* mutants. Our previous studies had shown that a deletion of the *nodI* and *nodJ* genes had no negative effect on the nodulation ability of *B. japonicum* on soybean (Göttfert *et al.* 1989). To generate *nodS* and *nodU* mutations, we inserted the Ω interposon into *nodS* and the kanamycin resistance gene into *nodU*, resulting in mutants Δ440 and 434 (Fig. 1 and Table 1). These mutants were tested for nodulation on soybean, cowpea, mung bean, Siratro, and *G. soja*. The result was that all plant species were nodulated similarly as they were by the wild type (data not shown).

Complementation of NGR234Δ1, a derivative of *Rhizobium* sp. strain NGR234, by *nodSU* of *B. japonicum*. NGR234Δ1, a deletion derivative of *Rhizobium* sp. strain NGR234, is severely reduced in its ability to nodulate *L. leucocephala* (Lewin *et al.* 1990). This defect could be complemented by *nodS* of *Rhizobium* sp. strain NGR234 (Lewin *et al.* 1990). Therefore, we mobilized pRJ453 (Fig. 1) carrying the *B. japonicum nodSU* genes expressed constitutively by the kanamycin resistance promoter into NGR234Δ1. However, *L. leucocephala* seedlings inoculated with NGR234Δ1 (pRJ453) failed to develop nodules even after 3 wk. Equivalent data were obtained by Lewin *et al.* (1990).

DISCUSSION

Structural organization of the *B. japonicum* common *nod* gene region. There are two striking features that emerged from the established DNA sequence of the *nodS*, *nodU*, *nodI*, and *nodJ* genes. The first one is the 71-bp long overlap of the *nodC* and *nodS* genes (Fig. 2). Because the overlap involved the 3′ end of *nodC*, we compared the carboxy-terminal ends of the *nodC* gene products of *B. japonicum*, *R. meliloti* (Török *et al.* 1984), and *R. l.* bv. *viciae* (Rossen *et al.* 1984) (data not shown). All three NodC proteins are highly conserved up to the arginine residue encoded by nucleotides 45–47 as shown in Figure 2. Thereafter, the amino acid sequences diverge: the NodC proteins of *R. meliloti* and *R. l.* bv. *viciae* continue for an additional 31 weakly conserved amino acid residues, whereas the NodC protein of *B. japonicum* extends for another 90 nonconserved amino acid residues and, thus, is much longer. To exclude the possibility that this extension (which includes the overlapping region) was due to a DNA sequencing error, we confirmed the nucleotide sequence of this region in two additional, independently isolated clones. The assumption that the *nodS* gene indeed starts way before *nodC* ends is also corroborated by the fact that the first 24 amino acids of the NodS protein encoded by the overlapping sequence are well-conserved (63%) in the amino-terminal region of the NodS protein of *Rhizobium* sp. strain NGR234 (Fig. 3).

Overlapping genes are often found in viral but less frequently in bacterial genomes. Much more common is the phenomenon called "translational coupling" in which two sequential bacterial genes overlap only with their translation termination (first gene) and translation initiation signals (second gene) (reviewed by Normark *et al.* 1983). Such a situation appears to be realized between *nodU* and *nodI* and between *nodI* and *nodJ* (Fig. 2). These short distances, including that between *nodS* and *nodU*, also suggest that all genes are located within the same operon. While translational coupling facilitates the initiation of translation of genes located farther downstream in a poly-

A

```
BjNodI   1      MNMSNMAIDLVGVRKSFGDKVIVNDLSFSVARGECFGLLGPNGAGKSTIARMLLGMISPDRGKITVLDEPVPSRARAARVAVGVV
RlNodI   1 MDSPSGSLSPVAIDLAGVSKSYGGKIVVNDLSFTIAAGECFGLLGPNGAGKSTITRMILGMTSPSVGKITVLGAQEPGQVRLARAKIGIV

BjNodI  86 PQFDNLEPEFTVRENLLVFGRYFGMSARTIEAVVPSLLEFARLESKADVRVSLLSGGMKRRLTLARALINDPHLLVMDEPTTGLDPHARH
RlNodI  91 SQFDNLDLEFTVRENLLVYGRYFRMSTREIETVIPSLLEFARLESKANTRVADLSGGMKRRLTLAGALINDPQLLILDEPTTGLDPHARH

BjNodI 176 LIWERLRALLARGKTILLTTHFMEEAERLCDRLCVLESGCKIAEGKPDALIDEHIGCNVIEIYGGDLDQLRELIRPYARHIEVSGETLFC
RlNodI 181 LIWERLRSLLARGKTILLTTHIMEEAERLCDRLCVLEAGRKIAEGRPHALIEEQIGCPVIEIYGGDPQELSLLIRPNARRLEISGETLFC

BjNodI 266 YARCPDEISVHLRGRTDLRVLQRPPNLEDVFLRLTGREMEK 306
RlNodI 271 YTPDPEQVRAQLRAYSNLRLLERPPNLEDVFLRLTGREMEK 311
```

B

```
BjNodJ   1 MDDGYASVMPANAYNWTAVWRRNYLAWRKVALASLLGNLADPITNLFGLGFGLGLIVGRVEGTSYIAFLAAGMVAISAMTSATFETLYAA
RlNodJ   1    MGVATLPAGGLNWLAVWRRNYLAWKKAALASILGNLADPVIYLFGLGAGLGVMVGRVDGVSYTAFLAAGMIATSAMTAATFETIYAA

BjNodJ  91 FARMDVKRTWEGILFTQLTLGDIVLGELVWAASKSVLAGTAIGIVAATLGYASWTSVLCAIPTIALTGLVFASLAMVVISLAPTYDYFVF
RlNodJ  88 FGRMQGQRTWEAMLYTQLTLGDIVVGEMAWAATKASLAGTGIGIVAAMLGYTHWLALLYALPVIAITGLAFASLGMVVTALAPSYDYFIF

BjNodJ 181 YQSLVLTPMVFLCGAVFPTSQMPDSFQHFAGLLPLAHSVDLIRPVMLERGADNAALHVGALCVYAVLPFFASIALFRRRLLR 262
RlNodJ 178 YQTLVITPMLFLSGAVFPVDQLPVAFQQIAAFLPLAHSIDLIRPTMLGQPIANVCLHIGVLCIYIVVPFLVSTALLRRRLMR 259
```

Fig. 5. Amino acid sequence alignment of the *Bradyrhizobium japonicum* NodI and NodJ proteins with the *Rhizobium leguminosarum* bv. *viciae* NodI and NodJ proteins. **A**, NodI comparison. Conserved amino acids of two motifs considered to be involved in ATP binding (Higgins *et al.* 1988) are underlined. The aspartate residue in the second conserved motif is always preceded by four hydrophobic amino acids (dotted line). **B**, NodJ comparison. Bj = *B. japonicum*; Rl = *R. l.* bv. *viciae*.

cistronic mRNA, one must question whether the extensive *nodC*/*nodS* overlap serves a similar function. At least the relatively poor expression of a *nodS'-'lacZ* fusion as compared to the *nodC'-'lacZ* fusion suggests a hindrance of *nodS* translation, even though *nodC* and *nodS* are on the same operon.

The second interesting feature is the unusual overall composition and organization of the common *nod* gene cluster of *B. japonicum* as compared to other rhizobia. In *R. l.* bv. *viciae*, *R. l.* bv. *trifolii*, and *R. meliloti*, the *nodABC* genes are followed directly by *nodIJ* (see Long 1989 for a recent review). In *B. japonicum*, the *nodABC* genes are preceded by *nodY* (Banfalvi *et al.* 1988) and followed by *nodSU* and then *nodIJ*. It is interesting to learn that *nodS* and *nodU* are also two adjacent genes in *Rhizobium* sp. strain NGR234, but separated from *nodC* (Lewin *et al.* 1990). It is possible, therefore, that the *nodYABCSUIJ* cluster in *B. japonicum* has evolved as a result of an insertion of *nodSU* between *nodC* and *nodI*.

Regulation of *nodC*, *nodS*, and *nodU* gene expression. Using translational *nodC'-'lacZ*, *nodS'-'lacZ*, and *nodU'-'lacZ* fusions, we could show that the expression of these genes was inducible by daidzein (Table 2), and demonstrated that their induction was dependent on the presence of an upstream DNA fragment carrying the *nod*-box and, hence, probably the promoter. It has to be emphasized that the *nodC'-'lacZ* and *nodS'-'lacZ* fusions were constructed within the same *Sal*I restriction site at nucleotide sequence position 292 (Fig. 2). This site is located in that part of the 3' end of *nodC* that has no equivalent in *nodC* of *R. meliloti* or *R. l.* bv. *viciae*. Thus, the inducer-dependent expression of *nodC*, *nodS*, and *nodU* proves that the ORFs of *nodC* and *nodS* overlap, that *nodC* of *B. japonicum* is indeed longer than the known *nodC* genes of the fast-growing strains of *Rhizobium*, and that *nodC*, *nodS*, and *nodU* are located within an operon. Although the *nodSU* genes of *Rhizobium* sp. strain NGR234 are on a separate transcription unit, their transcription is also controlled by an inducible *nod*-box promoter (Lewin *et al.* 1990). Obviously, there is a strict requirement in both species for tightly regulated expression of these genes together with the common *nod* genes.

In the case of the *B. japonicum nodS'-'lacZ* fusion, the daidzein-dependent expression of β-galactosidase activity is low (twofold to threefold increase) but reproducible and always significantly above background (Table 2). As discussed before, this might be due to an inefficient *nodS* translation caused by the fact that it has to initiate before the end of *nodC*. The *nodC'-'lacZ* and *nodU'-'lacZ* fusions produce some β-galactosidase activity even in the absence of daidzein, and this activity drops to background levels when the upstream DNA fragment containing the *nod*-box is deleted (compare Fig. 1 and Table 2). This seems to indicate that some, if not all, nodulation genes in the cluster are expressed to a low level even in the absence of the inducer. It must be emphasized, however, that the gene expression values were obtained with *nod'-'lacZ* fusions located on plasmids, whereas the natural location of the *B. japonicum* nodulation genes is on the chromosome.

Table 2. Activation of nodulation genes in *Bradyrhizobium japonicum* by daidzein

Plasmids	Fusions	Deletion of *Xho*I fragment[a]	Daidzein (1 μM)	β-Galactosidase activities[b] (Miller units)
pRJ457*[c]	*nodC/S'-'lacZ*		+[d]	2.2 ± 0.8
pRJ457*	*nodC/S'-'lacZ*		−	2.8 ± 0.4
pRJ458	*nodC'-'lacZ*		+	289.0 ± 19.0
pRJ458	*nodC'-'lacZ*		−	48.0 ± 9.0
pRJ459	*nodS'-'lacZ*		+	5.6 ± 1.7
pRJ459	*nodS'-'lacZ*		−	2.2 ± 0.4
pRJ462	*nodU'-'lacZ*		+	51.0 ± 7.2
pRJ462	*nodU'-'lacZ*		−	6.8 ± 1.1
pRJ461*	*nodU'-'lacZ*		+	2.0 ± 0.7
pRJ461*	*nodU'-'lacZ*		−	2.2 ± 0.4
pRJ463	*nodC'-'lacZ*	Δ	+	2.4 ± 0.5
pRJ463	*nodC'-'lacZ*	Δ	−	2.6 ± 0.5
pRJ464	*nodS'-'lacZ*	Δ	+	2.4 ± 0.5
pRJ464	*nodS'-'lacZ*	Δ	−	2.2 ± 0.4
pRJ465	*nodU'-'lacZ*	Δ	+	2.0 ± 0.0
pRJ465	*nodU'-'lacZ*	Δ	−	1.8 ± 0.4

[a] Removes the *nod*-box and adjacent DNA regions (Fig. 1).
[b] Mean values of five independent cultures. Decimals are given for values below 10.
[c] * = Plasmids containing out-of-frame *nod'-'lacZ* fusions that served as background controls.
[d] + = present; − = absent.

Phenotype of *nodS* and *nodU* mutant strains and potential function of the *nodS*, *nodU*, *nodI*, and *nodJ* gene products. It was somewhat disturbing that the *nodS* and *nodU* mutant strains did not show any significant reduction in nodulation ability on soybean and several other host plants. Moreover, this was unexpected because a *Rhizobium* sp. strain NGR234 *nodS* mutant showed considerable nodulation delay on cowpea and was even Nod⁻ on one of its hosts, *L. leucocephala* (Lewin *et al.* 1990). Interestingly, however, *L. leucocephala* has never been reported to be a host plant for *B. japonicum*, and with our assay conditions we also did not observe any nodulation of this plant by the *B. japonicum* wild type. Furthermore, efforts to complement a deletion derivative of NGR234, lacking the *nodS gene*, by the *nodSU* genes of *B. japonicum* failed. At present, the only conclusions we can offer are that the corresponding *B. japonicum* wild-type *nodSU* genes may be important for the nodulation of another, hitherto unknown host plant, or they may serve a function which is necessary under natural conditions in the soil or rhizosphere, that is in an environment that does not exist in our artificial test system.

If the *nodSU* genes were indeed determinants of host specificity, their location within the common *nod* gene cluster, though intriguing, is not without precedent. For example, in *R. l.* bv. *viciae* TOM an additional nodulation gene, *nodX*, a determinant of host specificity, is located downstream of *nodJ* and cotranscribed with the *nodABCIJ* operon (Davis *et al.* 1988). Despite the fact that we could not associate any observable phenotype with the *nodS* and *nodU* mutants, it seems unlikely that the NodS and NodU proteins of *B. japonicum* do not have any function at all. First, the *nodSU* genes appear to be well-conserved in many *Bradyrhizobium* species and strains (Fig. 4), and second, they are expressed and regulated in a coordinate way together with the other common *nod* genes located upstream of them.

The NodI and NodJ proteins were found to be highly homologous to the NodI and NodJ proteins of *R. l.* bv. *viciae* (Fig. 5). Furthermore, the NodI protein was found to be homologous to a family of ATP-binding proteins, a finding first made by Evans and Downie (1986). These ATP-binding proteins are constituents of osmotic shock-sensitive permeases, which are essential for the uptake of certain small, hydrophilic molecules across the cytoplasmic membrane. Thus, the *nodI* gene product together with the highly hydrophobic *nodJ* gene product might serve a function related to the transport of such a molecule.

ACKNOWLEDGMENTS

This work was supported by the Federal Institute of Technology, Zürich. M. Göttfert received a fellowship from the Deutsche Forschungsgemeinschaft.

We wish to thank A. Lewin and W. J. Broughton for communicating their results on strain NGR234.

LITERATURE CITED

Banfalvi, Z., Nieuwkoop, A., Schell, M., Besl, L., and Stacey, G. 1988. Regulation of *nod* gene expression in *Bradyrhizobium japonicum*. Mol. Gen. Genet. 214:420-424.

Davis, E. O., Evans, I. J., and Johnston, A. W. B. 1988. Identification of *nodX*, a gene that allows *Rhizobium leguminosarum* biovar *viciae* strain TOM to nodulate Afghanistan peas. Mol. Gen. Genet. 212:531-535.

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.

Evans, I. J., and Downie, J. A. 1986. The nodI gene product of Rhizobium leguminosarum is closely related to ATP-binding bacterial transport proteins; nucleotide sequence analysis of the nodS and nodJ genes. Gene 43:95-101.

Feinberg, A. P., and Vogelstein, B. 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 137:266-267.

Fischer, H. M., Alvarez-Morales, A., and Hennecke, H. 1986. The pleiotropic nature of symbiotic regulatory mutants: *Bradyrhizobium japonicum nifA* gene is involved in control of *nif* gene expression and formation of determinate symbiosis. EMBO J. 5:1165-1173.

Göttfert, M., Weber, J., and Hennecke, H. 1988. Induction of a *nodA-lacZ* fusion in *Bradyrhizobium japonicum* by an isoflavone. J. Plant Physiol. 132:394-397.

Göttfert, M., Lamb, J. W., Gasser, R., Semenza, J., and Hennecke, H. 1989. Mutational analysis of the *Bradyrhizobium japonicum* common *nod* genes and further *nod* box-linked genomic DNA regions. Mol. Gen. Genet. 215:407-415.

Hennecke, H., Günther, I., and Binder, F. 1982. A novel cloning vector for the direct selection of recombinant DNA in *E. coli*. Gene 19:231-234.

Higgins, C. F., Gallagher, M. P., Mimmack, M. L., and Pearce, S. R. 1988. A family of closely related ATP-binding subunits from prokaryotic and eukaryotic cells. BioEssays 8:111-116.

Kosslak, R. M., Bookland, R., Barkei, J., Paaren, H. E., and Appelbaum, E. R. 1987. Induction of *Bradyrhizobium japonicum* common *nod* genes by isoflavones isolated from *Glycine max*. Proc. Natl. Acad. Sci. USA 84:7428-7432.

Lamb, J. W., and Hennecke, H. 1986. In *Bradyrhizobium japonicum* the common nodulation genes, *nodABC*, are linked to *nifA* and *fixA*. Mol. Gen. Genet. 202:512-517.

Lamb, J. W., Hombrecher, G., and Johnston, A. W. B. 1982. Plasmid determined nodulation and nitrogen fixation abilities in *Rhizobium phaseoli*. Mol. Gen. Genet. 186:449-452.

Lewin, A. 1989. Gènes symbiotiques de *Rhizobium* impliqués dans la reconnaissance de la plante-hôte. Ph.D. thesis, University of Geneva, Switzerland.

Lewin, A., Rosenberg, C., Meyer, R. H., Wong, C. H., Nelson, L., Manen, J. F., Stanley, J., Dowling, D. N., Dénarié, J., and Broughton, W. J. 1987. Multiple host-specificity loci of the broad host-range *Rhizobium* sp. NGR234 selected using the widely compatible legume *Vigna unguiculata*. Plant Mol. Biol. 8:447-459.

Lewin, A., Cervantes, E., Chee-Hoong, W., and Broughton, W. J. 1990. *nodSU*, two new *nod* genes of the broad host range *Rhizobium* strain NGR234 encode host-specific nodulation of the tropical tree *Leucaena leucocephala*. Mol. Plant-Microbe Interact. 3:317-326.

Long, S. R. 1989. *Rhizobium*-legume nodulation: Life together in the underground. Cell 56:203-214.

Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101:20-78.

Miller, J. H. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 466 pp.

Minton, N. P. 1984. Improved plasmid vectors for the isolation of translational *lac* gene fusions. Gene 31:269-273.

Nieuwkoop, A. J., Banfalvi, Z., Deshmane, N., Gerhold, D., Schell, M. G., Sirotkin, K. M., and Stacey, G. 1987. A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. J. Bacteriol. 169:2631-2638.

Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, B., Lindberg, F. P., and Olsson, O. 1983. Overlapping genes. Annu. Rev. Genet. 17:499-525.

Norrander, J., Kempe, T., and Messing, J. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101-106.

Prentki, P., and Krisch, H. M. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene 29:303-313.

Regensburger, B., and Hennecke, H. 1983. RNA polymerase from *Rhizobium japonicum*. Arch. Microbiol. 135:103-109.

Rossen, L., Johnston, A. W. B., and Downie, J. A. 1984. DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAB* and *C* required for root hair curling. Nucleic Acids Res. 12:9497-9508.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schofield, P. R., and Watson, J. M. 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.

Scott, K. F. 1986. Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium* sp. *Parasponia*. Nucleic Acids Res. 14:2905-2919.

Shine, J., and Dalgarno, L. 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71:1342-1346.

Simon, R., Priefer, U., and Pühler, A. 1983. Vector plasmids for in-vivo and in-vitro manipulations of gram-negative bacteria. Pages 98-106 in: Molecular Genetics of the Bacteria-Plant Interaction. A. Pühler, ed. Springer-Verlag, Berlin.

Török, I., Kondorosi, E., Stepkowski, T., Posfai, J., and Kondorosi, A. 1984. Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nucleic Acids Res. 12:9509-9524.

Vincent, J. M. 1970. A manual for the practical study of the root-nodule bacteria. IBP Handbook 15. Blackwell Scientific Publications, Oxford. 164 pp.