Volume 14 Number 7 1986 Nucleic Acids Research

Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium* sp. *(Parasponia)*

Kieran F.Scott

Centre for Recombinant DNA Research, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra, ACT 2601, Australia

Received 13 February 1986; Accepted 13 March 1986

ABSTRACT

A nodulation locus from the broad-host-range, non-legume symbiont *Bradyrhizobium* sp. *(Parasponia)* strain ANU289, has been identified by hybridisation to cloned *Rhizobium trifolii* nodulation (*nod*) genes. Transfer of cloned ANU289 *nod* genes to *R.trifolii* nodulation-deficient mutants showed that the locus contains a functional homologue of the *R. trifolii nodD* gene. DNA sequence analysis revealed the presence of three additional genes *nodA*, *nodB* and *nodC* clustered adjacent to *nodD*. The four genes from ANU289 share substantial sequence homology with those characterised from narrow-host-range *Rhizobium* strains. A novel 700-bp sequence inserted between the *nodD* and *nodABC* genes encodes an open reading frame designated *nodK* and is oriented in the same direction as *nodABC*. *nodKABC* appear to be organized in a single transcriptional unit and *nodD* is oriented divergently to *nodKABC*. A 35-bp sequence containing the ribosome binding site for the *nodD* gene and an AT-rich core sequence has been identified by comparison with sequences from other *Rhizobium* strains and is likely to be implicated in the plant-mediated induction of nodulation gene expression.

INTRODUCTION

The nodulation of plants by *Rhizobium* or *Bradyrhizobium* species and the subsequent fixation of atmospheric nitrogen within these nodules is the result of the co-ordinated expression of both plant and bacterial genes (1,2). This symbiotic association is usually restricted to plant species within the family Leguminosae (3), however, certain strains of bacteria are capable of nodule formation and nitrogen fixation in symbiosis with the non-legume tree *Parasponia* (Ulmaceae) (4,5). One such organism *Bradyrhizobium* sp. *(Parasponia)* strain ANU289 (previously called *Parasponia Rhizobium* (6)) is also able to effectively nodulate a number of tropical legume species including the forage legume *Macroptilium atropurpureum* cv. siratro.

The mode of infection and nodule structure induced by ANU289 on each of these plants is morphologically quite different (7). Infection of siratro occurs through root hair curling and infection thread formation, resulting in a determinate nodule structure comparable to that produced by bacteria on

other tropical legumes (7,8). However, with the infection of *Parasponia*, root hair curling has not been observed (9,10). Bacteria enter the root at points of bacterium-induced meristematic activity (10) and invade cortical cells through the induction of infection threads (9,10). The nodule formed, resembles a modified lateral root with an apical meristematic zone and a central vascular system (7,9,10).

Genetic analysis of nodulation in the *Rhizobium*-legume symbiosis has identified nodulation (*nod*) genes localised on large symbiotic (Sym) plasmids (11,12,13). The Sym plasmid-encoded nodulation genes are clustered adjacent to the *nif* genes in these strains (11,12,13). In one such strain, *Rhizobium trifolii*, a single 14kb fragment of DNA is capable of conferring the ability to nodulate clover plants (*Trifolium repens*) on *Agrobacterium tumefaciens* (14). Hybridisation studies and cross-species complementation experiments have shown that at least some nodulation genes are conserved in a large number of *Rhizobium* species (15,16,17,18,19,20,21). Four of these conserved genes, *nodA*, *nodB*, *nodC* and *nodD* have recently been characterised by DNA sequence analysis in several *Rhizobium* strains (22,23,24,25,26).

The genetic determinants which encode non-legume nodulation are unknown. Marvel *et al.* (27) have shown that a strain, with similar symbiotic properties to ANU289, contains genes which can functionally complement mutations in the *R.meliloti nod* genes *nodABC* and will hybridise to *R.meliloti* fragments carrying these genes. In this paper, a locus containing the four conserved genes *nodABC* and *nodD* has been isolated from ANU289 by hybridisation to *R. trifolii nod* gene sequences. DNA sequence analysis of this conserved region has identified a previously unreported open reading frame linked to *nodABC*. Analysis of this sequence predicts that it constitutes a gene, and it is designated *nodK*. In addition, comparative analysis of untranslated regions with those from other *Rhizobium* strains has identified a consensus sequence between the *nodKABC* and *nodD* genes which may be involved in the plant-mediated induction of *nod* gene expression.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

Bradyrhizobium sp *(Parasponia)* strain ANU289 is a streptomycin-resistant derivative of strain CP283 (5) and was previously called *Parasponia Rhizobium* sp ANU289 (6). *Rhizobium trifolii* strains ANU851, ANU274, ANU249 and ANU277 are Tn5-induced, nodulation-deficient mutants of

Rhizobium trifolii ANU843 and were obtained from P.R. Schofield (28). Plasmid pRt587 contains a 14kb *Hin*dIII fragment encoding *Rhizobium trifolii* nodulation genes (14) and pRt572 is a 7.2kb *Eco*RI subclone from this region (18). Both plasmids were obtained from J.M. Watson. Bacteriophage banks were maintained in *E.coli* strain LE392. Subclones for restriction analysis experiments were constructed in the plasmid pBR328 (29) and recombinants were maintained in *E.coli* strain RR1. Subcloning for DNA sequence analysis was into the M13 phage vectors mp18 and mp19 (30) and maintained in *E.coli* strain JM107. Complementation experiments were carried out using the broad-host-range plasmid pRK290 (31) using the helper plasmid pRK2013 in *E.coli* strain HB101.

Molecular cloning and hybridisation procedures

Restriction endonucleases were obtained from Boehringer Mannheim, Amersham International or New England Biolabs and used in TA buffer (32) under conditions described previously (33). T4 DNA ligase was a gift from Anne Mackenzie. DNA polymerase I was obtained from Bresa Inc. and used in accordance with their instructions.

The isolation of λ DNA, construction of genomic libraries and hybridisation conditions were as described previously (34). Hybridisation probes were made by randomly primed synthesis on denatured template DNA (35). Plasmid DNA was isolated as previously described (36). Transfer of DNA to nitrocellulose was according to the method of Southern (37). DNA was recovered from agarose gels by electroelution from low-gelling-temperature agarose (SeaPlaque, Marine Colloids). Transformation procedures were as previously described (35).

Bacterial mating and plant assay

DNA sequences to be transferred to *Rhizobium* strains were sub-cloned into the broad-host-range cloning vector pRK290 (31) and transformed into *E.coli* HB101. Donor *E.coli* strains and the helper strain HB101 (pRK2013) were grown without shaking in 2ml Luria broth at 37°C overnight. *Rhizobium trifolii* recipient strains were grown on TY slopes (38) for 3 days at 30°C and cells resuspended in 2ml sterile water prior to mating. Triparental patch matings were set up using 100 µl of each culture plated on TY plates and incubated overnight at 30°C. Crosses were replica plated on to selective medium (TM (39) containing tetracycline at 4 µg ml^{-1} and kanamycin at 150 µg ml^{-1}). Transconjugants were purified twice on selective medium and assayed for nodulation of white clover (*Trifolium repens*) by the plate method of Rolfe *et al.* (39).

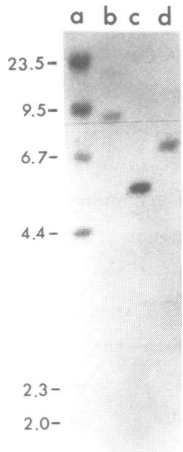

Figure 1: Identification of conserved nodulation locus in ANU289. DNA from a positively-hybridising phage λ PR289-4, isolated by hybridisation to *nod* gene sequences from *R.trifolii* ANU843 cloned in pRt587, (14) was prepared as described previously (34), cleaved with restriction enzymes and transferred to a nitrocellulose filter according to the method of Southern (37). The blot was hybridised at 65°C for 6 hours with [^{32}P]-labelled *R.trifolii nod* gene sequences isolated from pRt572 (17) by electroelution from agarose (Sea Plaque). (a) ^{32}P-labelled λ *Hin*dIII markers (sizes are in kb). (b) *Eco*RI. (c) *Bam*HI. (d) *Sal*I.

DNA Sequence Analysis

The source of DNA for all sequencing experiments was the plasmid pPR289-10. Restriction digests of this plasmid were electrophoresed on low-gelling-temperature agarose, fragments excised, ligated "in gel" into appropriately cleaved M13 vectors mp18 and mp19 and transformed into *E.coli* strain JM107 as previously described (33). The isolation and DNA sequencing of single-stranded template DNA was by the chain termination method as described previously (33). The M13 sequencing primer (5'-GTAAAACGACGGCCAGT-3) was constructed in our laboratory. DNA sequence was analysed using a modified version of the SEQ and ANALYSEQ programmes (40).

RESULTS

Identification of *Bradyrhizobium* sp. *(Parasponia) nod* genes

The Sym plasmid-encoded nodulation genes from the narrow-host-range *Rhizobium trifolii* strain ANU843 are located on a 14 kb *Hin*dIII fragment of DNA cloned in pRt587 (14). This 14 kb *Hin*dIII fragment was used as a hybridisation probe on a genomic library of ANU289 DNA, constructed in the lambda phage vector Charon 28 and several positively-hybridising clones were identified. DNA from one such clone, λ PR289-4, was isolated and the presence of sequences homologous to the *R.trifolii nod* probe was confirmed by hybridisation of various digests of λ PR289-4 with a 7.2 kb *Eco*RI fragment subcloned from pRt587. This fragment contains the *Rhizobium nod* genes which are conserved in other fast-growing organisms (18). As shown in Fig. 1, the 7.2 kb *Eco*RI probe hybridises to an 8.4 kb and 16.8 kb *Eco*RI fragment, a 5.3

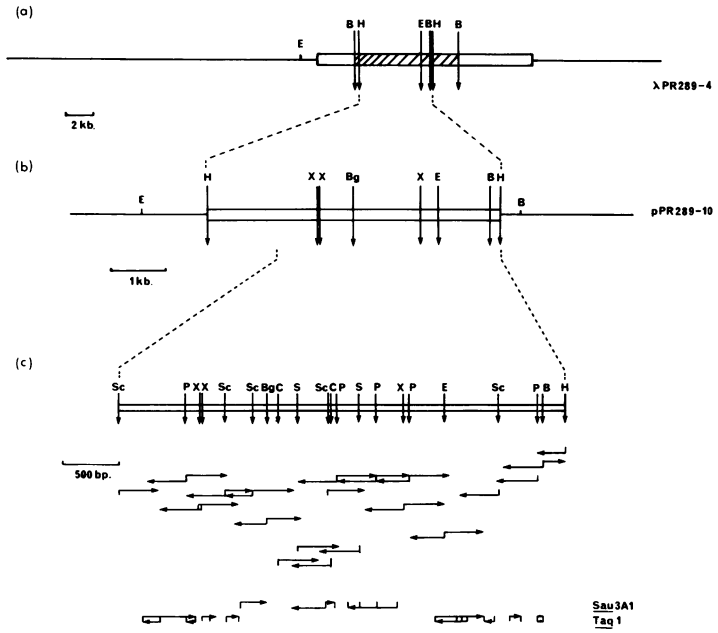

Figure 2: Isolation of ANU289 nodulation genes. (a) Restriction map of λ PR289-4. Shading indicates region of homology to *R.trifolii* nodulation gene probe cloned in pRT587 (14). Only those *Bam*HI sites referred to in the text are shown. (b) Restriction map of *nod* region cloned in pPR289-10, a 5.2 kb *Hin*dIII fragment cloned in pBR328. (c) DNA sequence strategy for ANU289 *nod* genes. Arrows indicate direction and length of sequence. Only those *Taq*I and *Sau*3AI sites used in sequencing are indicated. Restriction sites have been abbreviated as follows: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; X, *Xho*I; Bg, *Bgl*II; Sc, *Sac*I; P, *Pst*I; C, *Cla*I; S, *Sph*I.

kb *Bam*HI fragment and a 7.0 kb *Sal*I fragment from λ PR289-4. Further hybridisation analysis with the 14 kb *Hin*dIII probe (data not shown) localised the homology in this region to two contiguous *Bam*HI fragments in λPR289-4 as shown in Fig. 2a. Hybridisation of DNA from the entire λ PR289-4 clone to Southern blots of various pRt587 digests (data not shown) indicated that under the conditions used for these experiments, only sequences homologous to region I of the 14 kb *Hin*dIII fragment (28) could be detected in this locus. Hybridisation of λ PR289-4 DNA to genomic blots of ANU289 DNA (data not shown) confirmed the restriction map shown in Fig. 2a and showed that no reiteration of sequences cloned in λ PR289-4 is detectable in the genome.

Functional Complementation Studies

To determine if the cloned *nod* region of *Bradyrhizobium* sp.

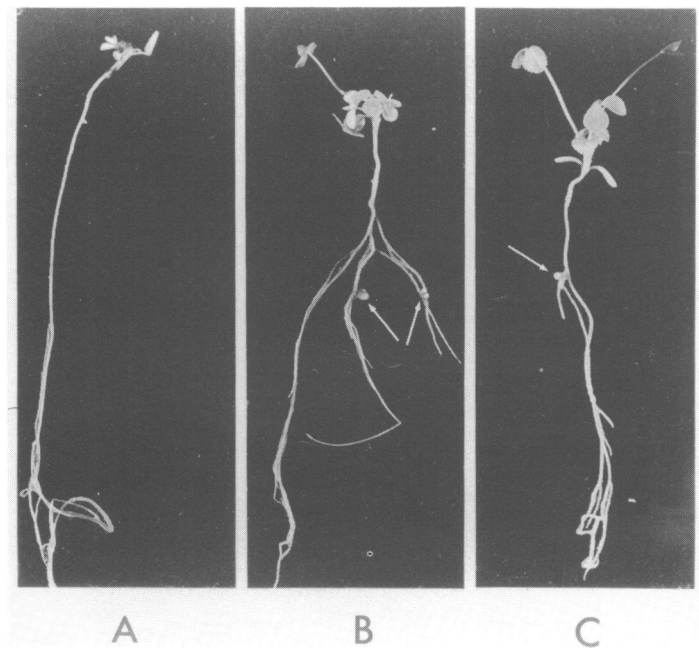

Figure 3: Functional complementation of *R.trifolii nodD* mutants. A 5.3 kb *Bam*HI fragment carrying the ANU289 nodulation locus cloned into pRK290 (pPR289-11) was transferred into *R.trifolii* nodulation-deficient mutants ANU274, ANU851, ANU249 and ANU277 as described. Tetracycline-resistant transconjugants were assayed for nodulation on clover plants (*Trifolium repens*) as previously described (39). A. mutant phenotype, ANU851, ANU274; B. ANU851 (pPR289-11); C. ANU274 (pPR289-11). Arrows indicate location of nodules.

(*Parasponia*) ANU289 encodes genes which are functionally homologous to *R.trifolii nod* genes, a 5.3 kb *Bam*HI fragment from λ PR289-4 (Fig. 2a), carrying most of the hybridising region, was cloned into the *Bgl*II site of the broad-host-range plasmid pRK290. This construct, pPR289-11, was then mated into the Tn5-induced, nodulation-deficient, *R.trifolii* mutants ANU851, ANU274, ANU249 and ANU277. Transconjugants from these crosses were assayed for nodulation of white clover (*Trifolium repens*). Of the four mutants tested, pPR289-11 was able to functionally complement only two strains, ANU851 and ANU274 (Fig. 3). Subsequent location of the position of Tn5 in these strains (26) indicated that the mutations in ANU851 and ANU274 are in the *nodD* gene of *R.trifolii* while the mutations in ANU249 and ANU277 are in *nodB* and *nodC*, respectively. These data indicate that pPR289-11 contains sequences which are functionally homologous to the *R.trifolii nodD* gene.

Figure 4: DNA sequence of ANU289 nodulation genes. Arrows indicate direction of translation. DNA sequences referred to in the text are indicated. SD Ribosome binding site (41).

Although it was possible to obtain nodules by complementation of *nodD* mutants with pPR289-11, the nodules induced on the plants by these transconjugants were much reduced in number (Fig. 3, B and C) and delayed in appearance by several weeks relative to the wild-type ANU843 (data not shown). Bacteria isolated from these nodules retained this delayed and reduced nodulation phenotype on white clover plants and also had the drug resistance and growth characteristics of the original transconjugants.

DNA sequence analysis

To determine the precise location and structure of the conserved nodulation genes in ANU289, a 5.3 kb *Hin*dIII fragment containing sequences homologous to the *R.trifolii* probe was subcloned from λ PR289-4 (Fig. 2a) into pBR328. This recombinant, pPR289-10, was mapped with a number of restriction endonucleases (Fig. 2b) and used for DNA sequence analysis. The complete DNA sequence of a 4.0 kb region of DNA from pPR289-10 was obtained by chain termination sequence analysis of various restriction fragments subcloned into appropriately cleaved M13 vectors mp18 and mp19 according to the strategy shown in Fig. 2c.

The DNA sequence of the conserved nodulation region in ANU289 is shown in Fig. 4. Six-phase translation of this sequence reveals five open reading frames (ORF's) greater than 400 codons in length. Comparative analysis of the translation products of these ORF's with available sequence data from narrow-host-range *Rhizobium* strains (22,23,24,25,26) has enabled the assignment of gene identity to four of these ORF's as shown in Fig. 4. This fragment carries the entire coding region of the *nodA*, *nodB* and *nodD* genes. In addition, the fragment encodes the N-terminal 153 codons of *nodC*. *nodA,B* and *C* are closely linked and oriented in the same direction while *nodD* is located 862 bp 5' to the initiation codon of *nodA* and is read divergently from *nodABC*.

Assignment of the initiation points of these genes is based on homology with homologous genes in the fast-growing rhizobia. In each case, the assigned initiation codons are the first available initiation codon in each ORF which provides maximum homology to the other three available sequences. The ORF's corresponding to *nodA*, *nodB* and *nodD* have no other potential initiation codons 5' to those assigned. The *nod*B initiation point has been assigned to a GTG codon overlapping the termination codon for *nod*A in an analagous position to the ATG initiation codons in the *R.meliloti* (22,24) and *R.trifolii* (26) sequences. No other potential initiation codons exist at the N-terminal end of this reading frame. The ORF corresponding to

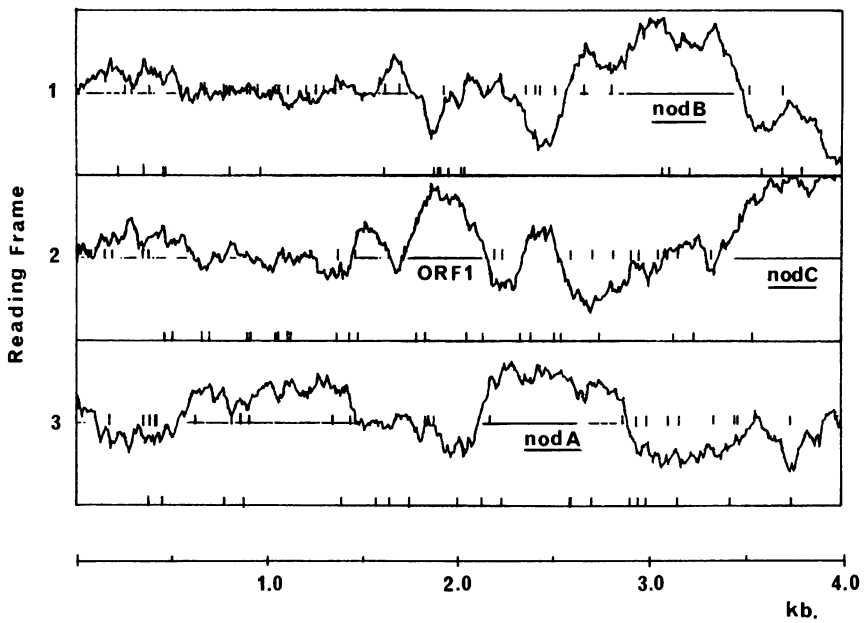

Figure 5: Identification of novel coding region (ORF1). DNA sequence (Fig. 4) was analysed using a positional base frequency programme modified from Staden (40). Relative position of known genes are indicated. Vertical lines on the baseline of each reading frame indicate the position of ATG codons. Vertical lines in the centre of each reading frame indicates the position of termination codons. Numbering of DNA sequence is as in Figure 4.

nodC extends 72 codons 5' to the assigned initiation codon, however the assigned ATG is the first conserved initiation codon in the ORF.

In addition to the four *nod* genes described above, this nodulation locus contains a novel open reading frame (ORF1) located 5' to the *nodABC* genes (Fig. 4). This reading frame is located within a 700-bp insertion of DNA between *nodD* and *nodABC* which is not present in the *R.trifolii*, *R.meliloti* or *R.leguminosarum nod* loci (22,23,24,25). Computer analysis of the potential for translation of this reading frame *in vivo* is shown in Fig. 5. From this analysis, it is clear that the programme is able to accurately identify the location of the *nodA*, *nodB* and *nodC* genes encoded on this strand. Therefore, by the criteria used to define translated regions in this programme (40), ORF1 is as likely to be translated as are *nodA*, *nodB* and *nodC*. This ORF is therefore tentatively assigned as *nodK*. Preceding the translational start point of the presumptive *nodK* gene there is a sequence 5'-ACTGGC-3' which is also found preceding the *nodC* gene

Table 1: Amino acid sequence conservation of nodulation gene products.

		R.leguminosarum	*R.meliloti*	*R.trifolii*
	nodA	42.9%*	57.1%	57.1%
	nodB	63.9%	60.2%	55.2%
B. sp. (*Parasponia*)				
	nodC	58.8%	62.7%	n.d.
	nodD	n.d.	61.2%	58.9%

*% homology is expressed as a ratio of the conserved amino acids in the two strains compared, as determined by the Wilbur-Lipman alignment procedure (48), upon the total number of amino acids in the *B.* sp. (*Parasponia*) sequence.
n.d. indicates insufficient data available to make the comparison.

(Fig. 4). This sequence does not resemble the usual purine rich ribosome binding sites of prokaryotic genes (41), however the *nodA* gene product is preceded by a more conventional ribosome binding site 5'-AGGGAAG-3' (Fig. 4).

The predicted size of the gene products for *nodA*, *nodB* and *nodD* (23.3 kD, 24.3 kD and 38.5 kD, respectively) are similar to those predicted for these genes in other organisms (22,23,24,26) while the presumptive *nodK* gene encodes a polypeptide of 14.9 kD. The degree of sequence homology between ANU289 and various *Rhizobium* strains for the *nodA*, *nodB* and *nodD* gene products and the N-terminal 153 codons of *nodC* is shown in Table 1. All four ANU289 gene products share between 55% and 64% sequence homology with the respective gene products from other *Rhizobium* strains, with the exception of the *R.leguminosarum nodA* protein. The low degree of homology (42.9%) is due to a truncation of the N-terminal and C-terminal regions of the protein in *R.leguminosarum* (23) relative to ANU289, *R.meliloti* (22,25) and *R.trifolii* (26).

Comparison of the untranslated regions of the ANU289 DNA sequence with available sequences from *R.meliloti* (24) and *R.trifolii* (26) has identified a consensus sequence 5' to the *nodD/nod(K)ABC* operons of the strains and preceding *nodFE* in *R.trifolii*. The degree of sequence conservation in this region is illustrated in Fig. 6. The consensus consists of an AT-rich core sequence of 23 nucleotides (5'-ATC$^{C}_{G}$AA$^{A}_{T}$CAATC$^{C}_{A}$ATTTTACCA-3') flanked by partially conserved sequences. The core sequence is located 33 bp 5' to the *nodD* initiation codon (Fig. 4) and the partial homology extends to within 12

```
B.p.  nodKABC  :-  ...GAGTCTATCCATCGTGTGGATGTATTCTATCGAAACAATCGATTTTACCAGATTGCGGA.....
R.t.  nodABC   :-  ...TTCTCGATCCACGCTGTAGATGATTGCGATCCAAACAATCAATTTTACCAATCTTTCGG.....
R.me. nodABC   :-  ...TGCGGCATCCATATCGCAGATGATCGTTATCCAAACAATCAATTTTACCAATCTTGCAG.....
R.t.  nodFE    :-  ...TCCTTCATCCATACTGCGGATGCTTTCGATCCAATCAATCAATTTTACCAATCCTTCGG.....

CONSENSUS      :-  (5').....ATCCA[T/C]NN[T/C][G/CA][TG]GATGN[ATTCT/TCGTG]ATC[G/C]AA[A/T]CAATC[G/A]ATTTTACCA[GATTG/ATCCT]N[GGA/CAG]...(3')
```

Figure 6: Identification of *nod* consensus sequence. Sequences were aligned for maximum homology using the Wilbur-Lipman alignment programme (48). B.p.; *Bradyrhizobium* sp. (*Parasponia*), R.t.; *Rhizobium trifolii* (26), R.me; *Rhizobium meliloti* (24).

bp of the *nodD* initiation codon in all strains.

DISCUSSION

Nodulation genes from the broad-host-range, non-legume nodulating strain *Bradyrhizobium* sp. (*Parasponia*) ANU289 have been identified and isolated by interspecific hybridisation to the cloned *nod* region of *R.trifolii* strain ANU843. DNA sequence analysis, together with functional complementation studies using *R.trifolii nod* mutants, show that this locus contains four genes *nodA*, *nodB*, *nodC* and *nodD* which are conserved in all *Rhizobium* strains examined to date (14,15,16,17,18,19,27). These genes are arranged in a similar fashion to the corresponding genes in *R.trifolii* (26), *R.meliloti* (22,25,26) and *R.leguminosarum* (23) with *nodABC* being contiguous and *nodD* located 5' to *nodABC* and read divergently.

The ANU289 *nod* locus contains a 700-bp insertion of DNA between *nodD* and *nodABC* which has not been located in any other *Rhizobium* strain to date. The insertion contains a novel open reading frame encoding a polypeptide of 14.9 kD which I tentatively designate *nodK*. Assignment of this ORF as a *nod* gene is based on the following evidence: (i) the ORF is located between a consensus sequence implicated in the expression of the *nodABC* operon in *R.trifolii* and *R.meliloti* and the *nodABC* genes of ANU289 indicating that the sequence is likely to be transcribed; (ii) computer analysis of potentially translated regions in the sequence (Fig. 5) predicts that ORF1 is as likely to be a translated coding region as are *nodA*, *nodB* and *nodC*; and (iii) a sequence 5'-ACTGGC-3' preceding the proposed initiation codon for the ORF is also located 5' to the *nodC* gene in an appropriate place to act as a ribosome binding site for translation. Although the role of this presumptive *nod* gene has not yet been determined,

it is of interest to note that preliminary hybridisation data indicates that this sequence has a structural homologue in *R.meliloti* strain Su47 and *R.leguminosarum* strain pRL1.

The close proximity of the *nodK*, *nodA*, *nodB* and *nodC* genes in this strain suggests that they constitute a single transcriptional unit. This view is supported by the failure of the clone pPR289-11 to complement the *R.trifolii nodB* mutant ANU249. The 5.3 kb *Bam*HI fragment cloned in pPR289-11 carries a complete *nodB* gene, but only the N-terminal 153 codons of *nodC*. The strong homology between the ANU289 *nodB* gene and the *R.trifolii nodB* gene, together with a previous report of complementation of *nodABC* mutants of *R.meliloti* by DNA from RP501, a related strain (27), indicates that complementation should occur. The most plausible explanation for the negative result is that the Tn5 insertion in ANU249 is polar on the expression of *nodC* which cannot be provided by pPR289-11 since it carries a truncated *nodC* gene. Similarly, the absence of a complete *nodC* gene on pPR289-11 explains its failure to complement the *nodC* mutant ANU277.

The ability of pPR289-11 to complement *nodD* mutants of *R.trifolii* was observed to be relatively inefficient resulting in a reduced and delayed nodulation phenotype on white clover. The stability of this delayed and reduced nodulation phenotype after passage through nodules and subsequent re-assay on clover plants indicates that the delay is not due to recombinational rescue of the mutant gene, as has been observed previously in *R. trifolii* complementation experiments (34), but rather may reflect impaired transcription and translation of the ANU289 gene in *R.trifolii*. Alternatively, the structural differences in the *nodD* gene product in the two strains may be sufficient to limit the functional activity of the ANU289 *nodD* protein in *R.trifolii*.

The translational start point of the *nodB* gene product overlaps the termination codons of *nodA* in this strain as in *R.trifolii* (26) and *R.meliloti* (22,24) indicating that translational coupling (42) of these two genes probably occurs. The significance of a requirement for translational control of expression of these two genes in particular is unclear.

The consensus sequence described in this paper and the accompanying paper (26) is of particular interest in light of recent reports on the regulation of expression of these genes (43,44,45). In *R.meliloti*, the expression of *nodABC* is enhanced by a root exudate and this enhancement requires the expression of *nodD* (43). Similarly in *R.trifolii* the expression of a number of nodulation genes, including *nodABC* and *nodFE*, is

inducible by exposure to clover plant roots (44). Thus, it seems likely that the observed consensus sequence is implicated in this plant-mediated induction of *nod* gene expression. By analogy with these systems, the presence of this sequence prior to the *nodKABC* operon in ANU289 implies that the expression of these genes is under a similar regulatory control. Given that ANU289 is capable of nodulation of several species of tropical legume and the non-legume *Parasponia* it will be of interest to determine if the expression of the *nodKABC* operon is obligatory for non-legume nodulation.

The proximity of the consensus sequence to *nodD* in the three strains strongly suggests that the sequence also influences the expression of the *nodD* gene. In fact, the 5' end of the consensus as shown in Fig. 6 incorporates the ribosome binding site for the *nodD* gene. The location of this sequence so close to the *nodD* gene and also preceding operons known to be inducible by root exudate suggests that it plays a bifunctional role in the regulation of *nod* gene expression. Search analysis of nucleotide sequence databases using the AT-rich core consensus sequence in both orientations indicates a high degree of homology between the core sequence as shown in Fig. 6 and the $PM_{II}7$ promoter recognised by a phage-modified RNA polymerase E gp28 in *Bacillus subtilis* (46). The mechanism of recognition of this promoter in *B. subtilis* is via the production of a phage-specific variant sigma factor which alters the specificity of the bacterial RNA polymerase. This observation suggests that the core sequence may be involved in transcriptional regulation of the *nodABC* genes.

The function of the *nodKABC* and *nodD* genes in ANU289 is unknown, although it seems likely that the *nodD* gene will have a similar regulatory role in the expression of *nodKABC* as in the *Rhizobium* strains. The *nodC* protein appears to be membrane located in *R. meliloti* (47) and will presumably be so in ANU289. The structure of *nodK*, *nodA* and *nodB* described here give no clues as to their potential function in legume or non-legume nodulation.

Since the mechanism of invasion and nodule differentiation by ANU289 in siratro and *Parasponia* appears to be quite different (7) it will be of interest to see if a common pathway of gene expression is required for the infection of both host plants. The isolation and characterisation of this locus in ANU289 will enable the construction of defined mutation in each of these genes and their regulatory regions to determine their role in the nodulation of non-leguminous plants.

ACKNOWLEDGEMENTS

I wish to thank Kim Farrant, Judy Sedger and Dr Heather Kane for excellent assistance with DNA sequencing; Dr Peter Schofield and Dr John Watson for provision of *R.trifolii* strains, and the plasmids pRt587 and pRt572; Anne Mackenzie for the gift of T4DNA ligase; David Smith for assistance with computer analysis of DNA sequences, Marilyn Olsen and Val Rawlings for the preparation of this manuscript and John Watson, Jeremy Weinman and Siiri Iismaa for helpful discussions. This work has been supported in part by an Agrigenetics-sponsored research grant to Dr. John Shine.

REFERENCES

1. Verma, D.P.S. and Long, S.L. (1983) International review of cytology suppl. 14. Jeon K.W., ed. (Academic Press, NY) pp211-245.
2. Rolfe, B.G. and Shine, J. (1984). In: Genes involved in Microbe-plant interaction. Verma, D.P.S. and Hohn, T.H., eds. (Springer-Verlag, Wien) pp95-128.
3. Vincent, J.M. (1974). In: Root-nodule symbiosis with *Rhizobium*. Quispel, A, ed. (North-Holland, Amsterdam) pp265-341.
4. Trinick, M.J. (1973). Nature 244:459-460.
5. Trinick, M.J. and Galbraith, J. (1980). New Phytol 86:17-26.
6. Weinman, J.J., Fellows, F.F., Gresshoff, P.M., Shine, J. and Scott, K.F. Nucl Acids Res 12:8329-8344.
7. Price, G.D., Mohapatra, S.S. and Gresshoff, P.M. (1984) Bot Gaz 145:444-451.
8. Ridge, R.W. and Rolfe, B.G. (1986). J Plant Physiol 122:121-137.
9. Trinick, M.J. and Galbraith, J. (1976) Arch Microbiol 108-159-166.
10. Lancelle, S.A. and Torrey, J.G. (1984) Protoplasma 123:26-37.
11. Rosenberg, C., Boistard, P., Denarie, J., Casse Delbart, F. (1981). Mol Gen Genet 194:326-333.
12. Beynon, J.L., Beringer, J.E. and Johnston, A.W.B. (1980). J Gen Microbiol 120:421-429.
13. Banfalvi, Z., Sakanyan, V., Koncz, C., Kiss, A., Dusha, I. and Kondorosi, A. (1981). Mol Gen Genet 184:318-325.
14. Schofield, P.R., Ridge, R.W., Rolfe, B.G., Shine, J. and Watson, J.M. (1984). Plant Mol Biol 3:3-11.
15. Long, S.R., Buikema, W. and Ausubel, F.M. (1982). Nature 298:485-488.
16. Hooykaas, P.J.J., Van Brussell, A.A.N., Den Dulk-Ras, H., Van Slogteren, G.M.S. and Schilperoot, R.A. (1981). Nature 291:351-354.
17. Kondorosi, E., Banfalvi, Z. and Kondorosi, A. (1984). Mol Gen Genet 193:445-452.
18. Djordjevic, M.A., Schofield, P.R., Ridge, R.W., Morrison, N.A., Bassam, B.J., Plazinski, J., Watson, J.M. and Rolfe, B.G. (1985). Plant Mol Biol 4:147-160.
19. Bachem, C.W.B., Kondorosi, E., Banfalvi, Z., Horvath, B., Kondorosi, A. and Schell, J. (1985). Mol Gen Genet 199:271-278.
20. Fischer, R.F., Tu, J.K. and Long, S.R. (1985). Appl Environ Microbiol 49:1432-1435.
21. Scott, D.B., Chua, K.Y., Jarvis, B.D.W. and Pankhurst, C.E. (1985). Mol Gen Genet 201:43-50.

22. Torok, I., Kondorosi, E., Stepkowski, T., Posfai, J. and Kondorosi, A. (1984). Nucl Acids Res 12:9509-9524.
23. Rossen, L., Johnston, A.W.B. and Downie, J.A. (1984). Nucl Acids Res 12:9497-9508.
24. Egelhoff, T.T., Fisher, R.F., Jacobs, T.W., Mulligan, J.T. and Long, S.R. (1985). DNA 4:241-248.
25. Jacobs, T.W., Egelhoff, T.T. and Long, S.R. (1985). J Bacteriol 162:469-476.
26. Schofield, P.R. and Watson, J.M. (1986) This issue.
27. Marvel, D.J., Kuldau, G., Hirsch, A., Richards, E., Torrey, J.G. and Ausubel, F.M. (1985). Proc Natl Acad Sci (USA) 82:5841-5845.
28. Djordjevic, M.A., Schofield, P.R. and Rolfe, B.G. (1985). Mol Gen Genet 200:463-471.
29. Soberon, X., Covarubias, L. and Bolivar, F. (1980). Gene 9:287-305.
30. Norrander, J., Kempe, T. and Messing, J. (1983). Gene 26:101-106.
31. Ditta, G., Stanfield, S., Corbin, D. and Helinski, D.R. (1980). Proc. Natl. Acad. Sci USA 77:7347-7351.
32. O'Farrell, P.H., Kutter, E. and Nakanishi, M. (1980). Mol Gen Genet 179:421-435.
33. Scott, K.F., Rolfe, B.G. and Shine, J. (1983). DNA 2:149-155.
34. Scott, K.F., Hughes, J.E., Gresshoff, P.M., Beringer, J.E., Rolfe, B.G. and Shine, J. (1982) J Mol Appl Genet 1:315-326.
35. Scott, K.F., Rolfe, B.G. and Shine, J. (1983). DNA 2:141-148.
36. Humphreys, G., Willshow, G.A. and Anderson, E.S. (1975). Biochem Biophys Acta 383:457-463.
37. Southern, E.M. (1975). J Mol Biol 98:503-517.
38. Beringer, J.E. (1974). J Gen Microbiol 84:188-198.
39. Rolfe, B.G., Gresshoff, P.M. and Shine, J. (1980) Plant Sci Lett 19:277-284.
40. Staden, R. (1984). Nucl Acids Res 12:551-567.
41. Shine, J. and Dalgarno, L. (1975). *Nature* 254:34-38.
42. Normark, S., Bergstrom, T., Jarvin, B., Lindberg, F.P. and Olsson, O. (1983). Ann Rev Genet 17:499-525.
43. Mulligan, J.T. and Long, S.R. (1985). Proc Natl Acad Sci USA 82:6609-6613.
44. Innes, R.W., Kuempel, P.L., Plazinski, J., Canter-Cremers, H., Rolfe, B.G. and Djordjevic, M.A. (1985). Mol Gen Genet (in press).
45. Rossen, L., Shearman, C.A., Johnson, A.W.B. and Downie, J.A. (1985). EMBO J (in press).
46. Greene, J.R., Brennan, S.M., Andrew, D.J., Thompson, C.C., Richards, S.H., Heinrikson, R.L. and Geiduschek, E.P. (1984). Proc Natl Acad Sci USA 81:7031-7035.
47. John, M., Schmidt, J., Wieneke, U., Kondorosi, E., Kondorosi, A. and Schell, J. (1985). EMBO J 4:2425-2430.
48. Wilbur, W.J. and Lipman, D.J. (1983). Proc Natl Acad Sci USA 80:726-730.