Volume 12 Number 24 1984 Nucleic Acids Research

DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAB* and *C* required for root hair curling

L.Rossen, A.W.B.Johnston and J.A.Downie*

John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

Received 27 September 1984; Revised and Accepted 6 November 1984

ABSTRACT

A 3.2kb fragment of DNA cloned from Rhizobium leguminosarum has been shown to contain the genes necessary for the induction of root hair curling, the first observed step in the infection of leguminous plants by R. leguminosarum. The DNA sequence of this region has been determined and three open reading frames were identified: genes corresponding to these open reading frames have been called nodA, nodB and nodC and are transcribed in that order. Mutations within the nodC gene completely blocked root hair curling. However, a subcloned fragment containing only the nodC gene did not induce normal root hair curling (although some branching was observed), indicating that the nodA and B genes may also be required for normal root hair curling. From an analysis of the predicted amino acid sequences of the nodAB and C genes it appeared unlikely that their products are secreted; therefore it is concluded that the induction of root hair curling could be due to a secreted metabolite.

INTRODUCTION

Bacteria of the genus Rhizobium have the special ability of infecting leguminous plants, and inducing the host to make root nodules within which the bacteria fix nitrogen. One feature of the symbiosis is that it is specific, particular legumes being nodulated only by certain Rhizobium species.

Despite the morphological complexity of the infection process, relatively few Rhizobium genes appear to be required for the determination of host range and for nodule development. At least in the fast-growing species, these nodulation genes appear to be clustered on large native symbiotic plasmids and are linked to genes required for nitrogen fixation (1,2,3). Thus in R. leguminosarum, which nodulates peas, the nodulation genes are located within a 10kb region of the symbiotic plasmid pRL1JI. This was shown by cloning this region and transferring it to a strain of R. phaseoli that had lost its symbiotic plasmid. The transconjugants could nodulate peas and nodule development appeared to be normal; root hair curling occurred, infection threads were induced and bacteroid forms of Rhizobium were found within the nodules (4). Within this 10kb region of nodulation DNA five classes of

mutants were obtained which were affected in nodulation. Class I mutants were deficient in nodulation and root hair curling; class II mutants did not form nodules but were delayed in root hair curling; class III mutants did not form nodules but did induce root hair deformation and mutants of classes IV and V were delayed and reduced in nodulation ability. The nodulation genes of R. trifolii and R. meliloti (which nodulate clover and alfalfa respectively) have also been cloned and nod mutants have been isolated and characterised (3,5,6).

A genetic complementation analysis of nod genes is difficult since (a) all of the defined nod mutants are transposon-induced or due to deletions, and (b) there is very strong selection by the plant for nodulation proficient revertants or recombinants. As an alternative we have sequenced R. leguminosarum nod genes. The DNA sequence of a 3 kb region of nodulation DNA, affected by the class I and class III mutations, has been determined and three genes were identified corresponding to three open reading frames. Independently the DNA sequence of a corresponding region of R. meliloti nodulation DNA was also established (7 and S.R. Long and T. Jacobs, personal communication), allowing a detailed comparison of the two regions of DNA in these Rhizobium species.

MATERIALS AND METHODS

Bacterial strains, bacteriophage and plasmids are listed in Table I. Media and growth conditions for Rhizobium and E. coli were as described by Beringer (8). Bacterial crosses were performed as described previously (9). Propagation of bacteriophage M13 was done as described by Sanger et al. (10).

Roots of Vicia hirsuta were examined for root hair curling by light microscopy essentially as described by Fahraeus (11). This plant is a small seeded legume which is nodulated normally by R. leguminosarum and is more convenient for microscopical analysis of early stages of infection than Pisum sativum (peas).

Restriction enzymes, DNA polymerase (large fragment) and DNA ligase were purchased from BRL, (U.K. Ltd) and used according to the manufacturers instructions. Transformation of E. coli, plasmid isolation, agarose and acrylamide gel electrophoresis, nick translations and DNA hybridisations were carried out essentially as described previously (1).

For DNA sequencing the 6.6 kb EcoRI fragment of DNA carrying nod genes (fig. 1) was isolated from an agarose gel and circularized using DNA ligase. The DNA was randomly broken by sonication and the ends were repaired using T4 DNA polymerase (P.L. Biochemicals). Fragments of 200-600 bp in size were isolated following size-fractionation by electrophoresis and ligated into

TABLE I

Bacterial Strains		Reference
E. coli ED8767		(19)
E. coli JM101		(10)
Rhizobium 8400	Strain of Rhizobium phaseoli cured of its symbiotic plasmid	(20)
Bacteriophage		
M13 mp8	Derivative of M13	(21)
Plasmids		
pRL1JI	R. leguminosarum native symbiotic plasmid	(1)
pRL1JI nodC128::Tn5	Derivative of pRL1JI	(15)
pRL1JI nodB42::Tn5	" " "	(4)
pIJ1246	Cloned 6.6kb EcoRI fragment containing nod genes	(15)
pKT230	Wide host range vector	(21)
pIJ1389	3.2kb BglII-EcoRI fragment containing nodABC cloned in pKT230	This work
pIJ1419	XmaIII-EcoRI fragment containing nodC cloned in the HpaI site of pKT230	This work
pIJ1420	As pIJ1419 but opposite orientation	This work

SmaI-cut M13 mp8 as described by Bankier and Barrell (12). Approximately 700 clones were screened for relevant inserts by DNA hybridization using the 3.2 kb BglII - EcoRI fragment cloned in pIJ1389 (fig. 1). M13 template DNA (1ul) was mixed with 1 ul phosphate buffer and spotted onto GeneScreen (New England Nuclear) and used in routine DNA dot-blot hybridization experiments. The DNA from the positively reacting clones was sequenced using the dideoxynucleotide chain termination method (10,13) and the samples were run on 40cm 6% polyacrylamide buffer gradient gels (12).

The DNA sequence was determined by sequencing each strand at least twice. The data were analysed by the DNA analysis programmes described previously (14).

RESULTS

Root hair curling genes.

Three classes of R. leguminosarum nod mutations had been shown to affect root hair curling: in mutants of classes I and III, root hair curling was absent (although class III mutants could deform root hairs) but in class II

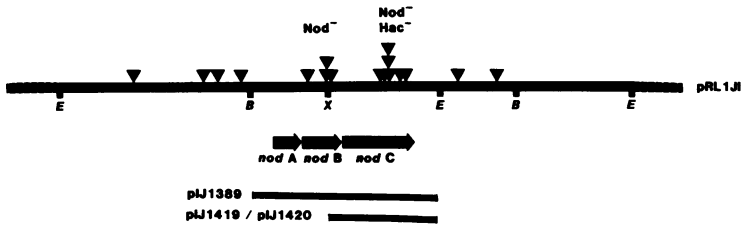


Fig. 1: R. leguminosarum nodulation region. The location of several previously described (15) nod::Tn5 alleles are shown by black triangles. The mutations in the nodABC region are all nodulation deficient (Nod⁻). The mutations in nodC are all totally defective in root hair curling or deformation (Hac⁻) whereas mutations in nodB do induce some root hair deformation. The restriction endonuclease sites shown are E, EcoRI; S, SstI; B, BglII; X, XmaIII. The nodA, nodB and nodC genes are indicated by arrows, and the regions cloned in pIJ1389, pIJ1419 and pIJ1420 are indicated.

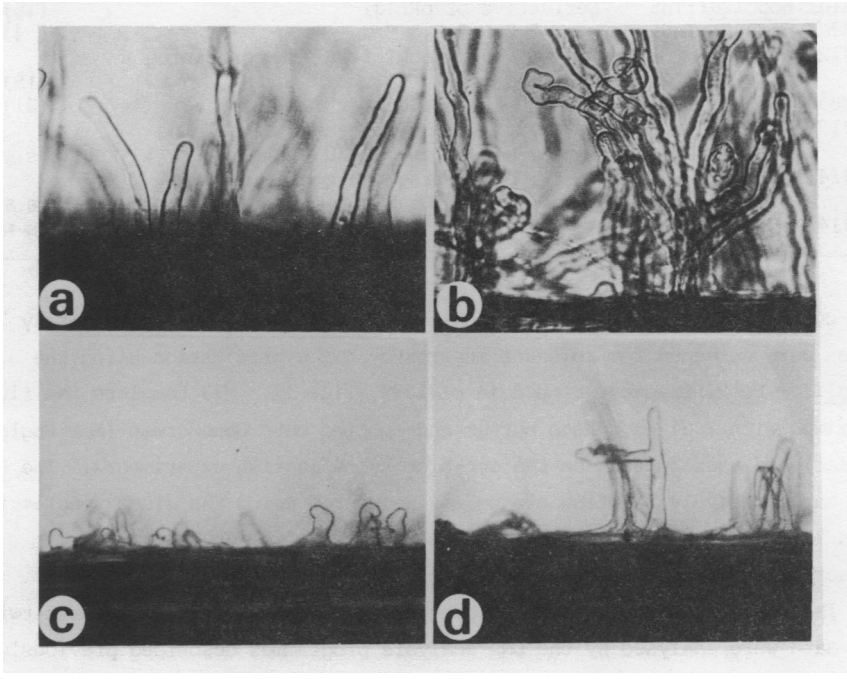


Fig. 2: Light micrographs of root hairs. The roots of Vicia hirsuta were examined under Nomarsky optics 7-10 days after inoculation. In (a) no root hair curling was observed with strain 8400, and a similar result was observed with strain 8400pRL1JI(nodC128::Tn5). In (b) strain 8400 was complemented for root hair curling by pIJ1389 and in (c) strain 8400 pRL1JI(nodC128::Tn5) was complemented for root hair curling by pIJ1420. In (d) occasional branched root hairs were observed with strain 8400pIJ1419.

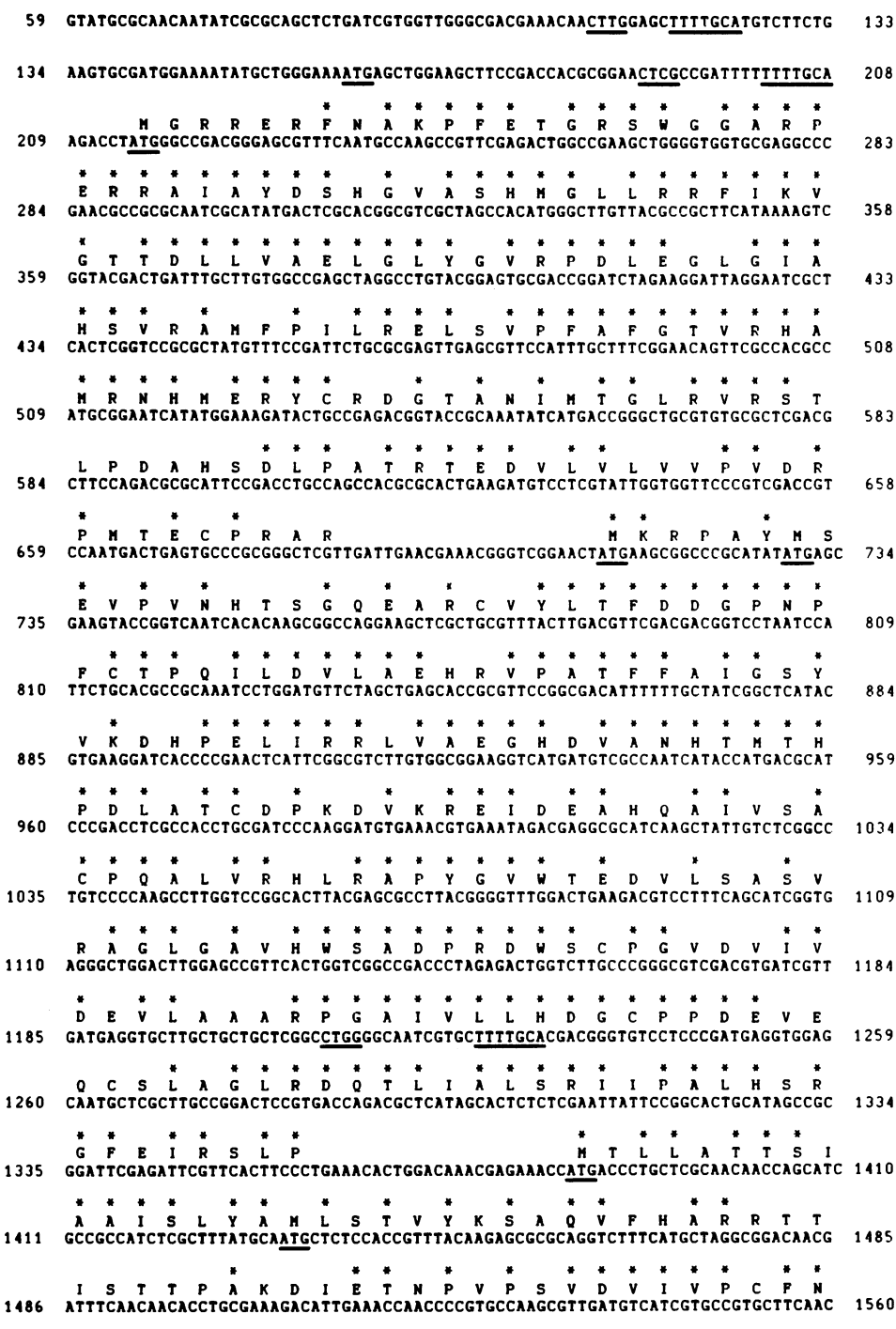
mutants root hair curling was delayed (15). The 3.2 kb EcoRI-BglII fragment of DNA containing the class I and III regions (see Fig 1) was subcloned into the EcoRI-BamHI sites of pKT230. The resultant plasmid pIJ1389, conferred upon a Rhizobium strain deleted for its nod genes the ability to induce root hair curling (Fig. 2a and b). Therefore the genes essential for root hair curling are contained within this region of DNA.

DNA sequence of nod genes.

The region of DNA corresponding to that cloned on pIJ1389 was sequenced using a series of randomly generated fragments of DNA cloned in M13 mp8. Within the sequenced region (Fig. 3) three open reading frames were identified (Figs. 1 and 3) all of which were in the same orientation. The genes corresponding to these open reading frames were called nodA, nodB and nodC (Fig. 1); the nodC gene corresponds to the region affected by class I mutants and the nodA and B genes correspond to the class III mutant region (Fig 1).

In an earlier study (15) which used an E. coli in vitro transcription/translation system it was shown that the nodC128::Tn5 allele (class I type) abolished the synthesis of a product of molecular weight about 45,000. Using plasmids pIJ1389 and pIJ1419 in an E. coli transcription/translation system it was confirmed that the nodC gene product has a molecular weight of about 45,000 (data not shown). As shown in Fig. 3 there are two potential ATG translation initiation sites (at positions 1384 and 1432, Fig. 3) for the nodC gene, and these would give products of predicted molecular weights 46,256 or 44,634. From these data it is unclear which translation initiation site is used although by comparison with the R. meliloti nodC gene (7) it would appear more likely that the first of the two translation initiation sites is used since only that ATG is common to both sequences.

No proteins were previously identified as corresponding to products of the nodA or nodB genes (15). In a further attempt to identify polypeptides specified by these genes, the in vitro translation products of pIJ1389 were examined, particularly in the molecular weight range 10,000-25,000. However, no new products corresponding to these genes were found. As seen in Fig. 3, there are two potential translation initiation codons for both the nodA (at 161 and 215) and the nodB (at 711 and 729) genes. These would give products of molecular weights 19,711 or 17,580 for the nodA gene and the predicted molecular weights of the nodB gene products would be 23,632 or 22,885. In the absence of protein analysis data it is difficult to know which of these starts are used. However with the nodA gene the second translation start would appear to be preferrable to the first because, between these two ATG codons



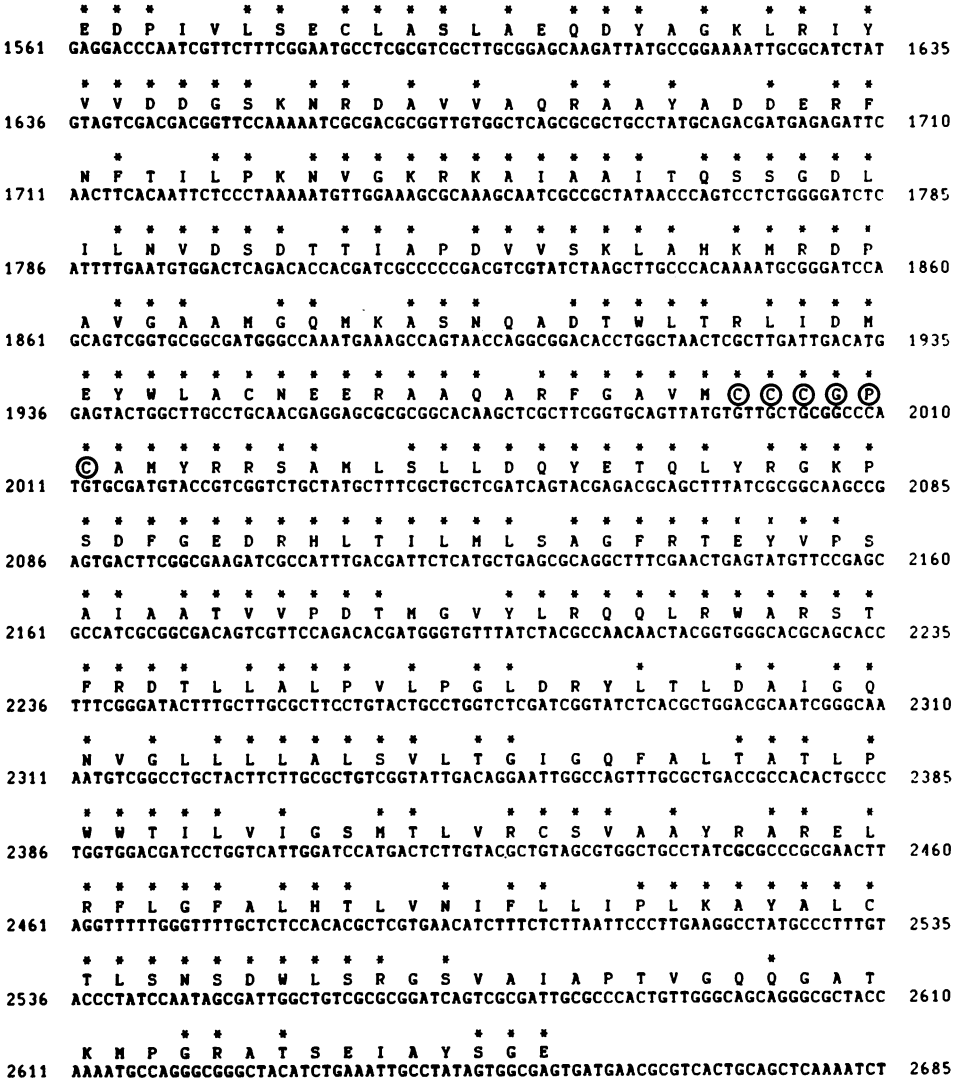


Fig. 3: R. leguminosarum nodA, nodB and nodC nucleotide and predicted amino-acid sequences. The amino-acids are indicated by the single letter code and those residues present in both R. meliloti and R. leguminosarum are asterisked, there appear to be two amino-acids deleted from the nodB gene of R. meliloti and one deleted from the nodC gene of R. leguminosarum (see Results). Potential ATG translation initiation codons are underlined as are promoter-like sequences (see Discussion). The unusual sequence CCCGPC circled affected the computer analysis of codon usage (see Results).

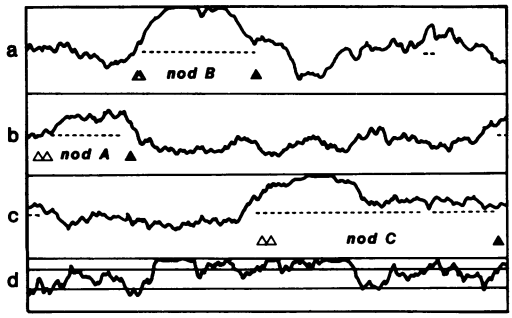


Fig. 4: Computer-assisted gene search. The diagrams (a), (b), (c) and (d) are the results of the DNA analysis programme ANALYSEQ (14) which search for genes on the basis of unequal codon use in protein sequences. In (a), (b) and (c) the effects of coding on positional base frequencies (using a window of 67 nucleotides) are shown for each of the three reading frames, and the most probable reading frames are indicated by broken lines (nodA, nodB, nodC). Appropriate initiation (open triangles) and termination (closed triangles) codons are indicated. In (d) Fickett's (23) method was used (with a window of 67 nucleotides) to asses the overall probability of the DNA sequence to contain coding regions. The output is plotted relative to three levels of decision: the top division indicates a coding region, the middle division is "no opinion" and the bottom division indicates non-coding regions.

the predicted amino-acid sequence contains a rather high proportion of infrequently used residues (such as tryptophane, phenylalanine and proline). Moreover, the computer-assisted analysis of the sequence (Fig. 4) indicated that this short region was unlikely to contain a reading frame. Accordingly, only the predicted amino-acid sequence downstream of the second ATG (at position 215) is shown.

Since no gene products corresponding to the nodA and nodB genes were identified the DNA sequence was analysed by two DNA sequence analysis programmes (14) which screen for gene coding sequences on the basis of a statistical analysis of codon usage. These analyses (Fig. 4) indicated that non-random sequences were present. Taken together with the results of another programme ("FRAME" ref 16, not shown) these analyses indicated that the non-random regions correspond to the three long open reading frames predicted from the initiation and termination codons. Near the centre of the nodC gene there is a region of DNA which has a low probability of being a protein coding sequence. However, this region corresponds to the 18 nucleotides coding for the amino acid sequence CCCGPC circled in Fig. 3. When these 18 nucleotides were deleted and the sequence re-analysed the drop in probability did not occur, and thus presumably is due to the infrequent codon usage for the cysteine residues.

The nodA, nodB and nodC genes are required for root hair curling.

As shown in Fig. 2, the DNA fragment (in pIJ1389) containing the nodABC genes can induce root hair curling. Previous studies (15) indicated that mutations of the nodC type (e.g. nodC128::Tn5) completely blocked root hair curling, whereas mutant strains carrying e.g. the nodB42::Tn5 allele had a partial effect, i.e. root hair deformation occurred. In order to ascertain if the nodC gene alone could induce root hair curling, the XmaIII-EcoRI DNA fragment (Fig. 1) containing nodC was subcloned (following end repair) into the HpaI site of pKT230 to form plasmids pIJ1419 and pIJ1420. The orientation of the inserted fragment in pIJ1420 is such that it is unlikely to be expressed strongly from a vector promoter whereas in pIJ1419 it is likely that the nodC gene is expressed from a vector promoter. The plasmids were then transferred to two Rhizobium strains: 8400 which is deleted for its nodulation genes and strain 8400 carrying the nodC128::Tn5 allele in the symbiotic plasmid pRL1JI. Plasmids pIJ1419 and pIJ1420 both corrected the nodC128::Tn5 allele for root hair curling (Fig. 2c). However with strain 8400 containing either plasmid, root hair curling was not observed (Fig. 2d), although a low frequency of branching of root hairs was observed. Therefore, since strain 8400 pIJ1389 containing the cloned nodAB and C genes can induce root hair curling, but the cloned nodC gene alone cannot, it appears that the nodA and/or nodB genes are required in addition to the nodC gene for root hair curling.

Comparison of the nodABC genes of R. leguminosarum and R. meliloti

Independently, the DNA sequence of three nod genes from R. meliloti (which nodulates alfalfa) has been established (7). Comparison of the two sets of DNA sequences shows that the three R. meliloti genes correspond to the R. leguminosarum genes described here and that the three genes are arranged in the same relative order. The amino-acids conserved in both species are asterisked in Fig. 3. The absolute amino-acid homology is about 70% and the DNA homology is about 70%. The proposed translational start of the R. meliloti nodA gene is 30 amino acids upstream of the second ATG (at position 215) in the R. leguminosarum nodA gene, but it should be noted that there is considerable DNA homology in this region.

Whereas in R. meliloti, the nodA and nodB genes appear to overlap by 4 nucleotides, this does not appear to occur in R. leguminosarum in which the nodA and B genes are separated by 27 nucleotides. Interestingly there is a high degree of conservation of these 27 nucleotides within the coding region of the R. meliloti nodA gene as indicated in the accompanying paper (7).

The nodB gene of R. leguminosarum contains two extra amino-acids (glutamine, Q and cysteine, C) at position 1260-1265 compared with the R. meliloti nodB gene. This addition occurs within a region which is otherwise highly conserved, but the significance (if any) of the difference remains to be established. Conversely, the R. meliloti nodC gene contains an extra amino-acid (glutamine, Q) which is absent (at position 1752) from the R. leguminosarum nodC gene.

DISCUSSION

The initially observed event in the infection of leguminous plants by Rhizobium is the induction of root hair curling. The DNA sequence of the R. leguminosarum (and R. meliloti, ref. 7) genes involved in this step of the symbiosis has been established and three genes, called nodAB and C, have been identified. These genes do not appear to be involved in the determination of the type of host plant nodulated, since the cloned R. leguminosarum nod genes could complement strains of R. meliloti, containing mutations in this region for the ability to nodulate alfalfa (3 and A. Kondorosi, personal communication).

It is unclear why the nodA and nodB gene products were not seen in the in vitro transcription/translation system: both predicted gene products contain methionine and thus should have been labelled with the S^{35}-methionine used, although the nodB gene has few methionine residues. The computer-assisted analysis of the DNA sequence (Fig. 4) corroborated that the open reading frames chosen (Fig. 3) corresponded to the most probable coding regions. In R. meliloti, gene products of molecular weights abot 23,000 and 28,500 were identified (24) and these may correspond to products of the nodA and B genes from R. meliloti (7).

Mutations affecting the nodC gene completely abolish root hair curling or deformation, but strains containing Tn5 within the nodB gene retained the ability to induce some root hair deformation or branching. Since Tn5 induces polar mutations and the nodC gene is downstream of the nodB gene, it would therefore appear that the nodC gene has its own promoter.

Interestingly, the nodC gene cloned in either orientation (in pIJ1419 and pIJ1420) could complement nodC128::Tn5 for root hair curling. Plasmid pIJ1420 (which contains the nodC gene in an orientation unlikely to be expressed from a vector promoter; ref 22 and unpublished observations) and pIJ1419 conferred upon strain 8400 the ability to induce some branching of root hairs. These observations also support the hypothesis that the nodC gene may have its own

promoter.

The DNA sequence CTGGNNNNNNTTTTGCA has been observed at -26 to -10 nucleotides preceeding the initiation of transcription of a number of R. meliloti nitrogen fixation genes (17). A similar sequence (at position 1212-1233 and contained within pIJ1419 and pIJ1420) preceeds the nodC gene, although it should be noted that this sequence occurs within the nodB gene coding region and the spacing between the conserved nucleotides is different. Moreover, two somewhat similar sequences (at positions 110-124 and 190-208) preceed the nodA gene. These sequences have been underlined in Fig. 3; it can be seen that when they are compared with each other, further similarities can be found in the adjacent regions. However, the significance of these regions as possible promotors has yet to be established.

Little is known about the biochemical roles of the nodA, nodB and nodC gene products. An examination of the amino-terminal regions of the predicted amino-acid sequences indicates that they are unlikely to contain transit peptides. Such transit peptides normally contain a short stretch of two to eight amino-acids containing positively-charged residues followed by a stretch of 15-18 uncharged hydrophobic amino-acids. None of the potential amino-terminal regions of nodA, nodB or nodC gene products conform to this pattern and the gene products are therefore presumably not exported. The carboxy-terminal region of the nodC gene product is somewhat hydrophobic (our observatins, data not shown) as is true of the R. meliloti nodC gene product (7 and S.R. Long and T. Jacobs, personal communication) suggesting that it may be membrane - associated. Since the nodA, nodB and nodC gene products are themselves unlikely to be exported and yet a strain containing pIJ1389 carrying only nodAB and C genes can induce root hair curling, it would appear then that these genes may be making (or modifying) some metabolic intermediate which could be secreted and then sensed by the plant root hair. The involvement of such secreted metabolites in nodulation has been postulated (18) although no conclusive proof for their existence has yet been presented.

ACKNOWLEDGEMENTS

We are indebted to C.D. Knight for help and assistance given on the analysis of root hair curling and growth of V. hirsuta. We thank I. Torok, A. and E. Kondorosi and S.R. Long, and T. Jacobs and their colleagues for making available DNA sequence data prior to publication.

We thank J. Walker and V. Tybulevicz for assistance in the construction of the random M13 clone bank and for advice on DNA sequencing and M.J. Bibb

for advice on DNA sequence analysis and A. G. Hepburn for help with computing. I. Evans and D. A. Hopwood critically read the manuscript. L.R. was supported by a Scientific and Technical Training Grant from the Commission of the European Communities and Danish Agricultural and Veterinary Research Council. Additional support came from EEC contract number GBI-5-070-UK.

*To whom correspondence should be addressed

REFERENCES

1 Downie, J.A., Ma, Q-S., Knight, C.D., Hombrecher, G. and Johnston, A.W.B. (1983) EMBO J. 2, 947-952.
2 Schofield, P.R., Djordjevic, M.A., Rolfe, B.G., Shine, J. and Watson, J. (1983) Mol. gen. Genet. 192, 459-465.
3 Kondorosi, E., Banfalvi, Z. and Kondorosi, A. (1984) Mol. gen. Genet. 193, 445-452.
4 Downie, J.A., Hombrecher, G., Ma, Q-S., Knight, C.D., Wells, B. and Johnston, A.W.B. (1983) Mol. gen. Genet. 190, 359-365.
5 Schofield, P.R., Ridge, R.W., Rolfe, B.G., Shine, J. and Watson, J. (1984) Plant Mol. Biol. 3, 3-11.
6 Long, S.R., Buikema, W.J. and Ausubel, G.M. (1982) Nature 198, 485-489.
7 Torok, I., Kondorosi, E., Stepkowski, T., Posfai, J. and Kondorosi, A. (1984) Nucl. Acid Res. this issue.
8 Beringer, J.E. (1974) J. gen. Microbiol. 87, 188-198.
9 Buchanan-Wollaston, A.V., Beringer, J.E., Brewin, N.J., Hirsch, P.R. and Johnston, A.W.B. (1980) Molec. gen. Genet. 178, 185-190.
10 Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci., USA 74, 5463-5476.
11 Fahraeus, G. (1957) J. gen. Microbiol. 16, 374-381.
12 Bankier, A.T. and Barrell, B.G. (1983) in The Life Sciences, vol. B5, Flavell, R.A. Ed. Elsevier (Ireland).
13 Sanger, F., Coulsen, A.R. Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. mol. Biol. 162, 729-773.
14 Staden, R. (1984) Nucl. Acid Res. 12, 551-567.
15 Downie, J.A., Knight, C.D., Johnston, A.W.B. and Rossen, L. (1984) Mol. Gen. Genet., in press.
16 Bibb, M.J., Findlay, R.P. and Johnston, M.W. (1984) Gene in press.
17 Better, M., Lewis, B., Corbin, D., Ditta, G. and Helinski, D.R. (1983) Cell 35, 479-485.
18 Syono, K., Newcomb, W. and Torrey, J.G. (1976) Can. J. Bot. 54, 2155-2162.
19 Murray, N.E., Brammar, W.J. and Murray, K. (1977) Mol. gen. Genet. 150, 53-61.
20 Lamb, J.W., Hombrecher, G. and Johnston, A.W.B. (1982) Mol. gen. Genet. 187, 166-171.
21 Messing, J., and Vieira, J. (1982) Gene 19, 269-276.
22 Bagdasarian, M., Lurz, R., Duckert, B., Franklin, F.C.H., Bagdasarian, M.M., Frey, J. and Timmis, K.N. (1981) Gene 16, 237-247.
23 Fickett, J.W. (1982) Nucl. Acid Res. 10, 5303-5318.
24 Schmidt, J., John, M., Kondorosi, E., Kondorosi, A., Wienecke, U., Schröder, G., Schröder, J., and Schell, J. (1984) EMBO J. 3, 1705-1711.