Genetic and physical analysis of the *nodD3* region of *Rhizobium meliloti*

Brenda G.Rushing, M.Melanie Yelton and Sharon R.Long*
Stanford University, Department of Biological Sciences, Stanford, CA 94305-5020, USA

Received September 13, 1990; Revised and Accepted January 15, 1991 EMBL accession no. X53820

ABSTRACT

The nodulation (*nod*) genes of the symbiont *Rhizobium meliloti* are transcriptionally controlled by protein activators in the *nodD* gene family. While NodD1 and NodD2 act in concert with small molecular weight inducers provided by the host legume plant, NodD3 is an inducer-independent activator of the *nod* promoters. We determined the sequence of the *nodD3* gene, confirmed the expression of a 35 kDa protein *in vitro*, and determined the insertion points of five Tn5 insertions in the region of the *nodD3* gene. We found the NodD3 amino acid sequence to be markedly diverged from the sequences of NodD1 and NodD2, which were more similar to the inducer-dependent NodD of another species, *Rhizobium leguminosarum* biovar *viciae*. The expression of *nodD3* is not well understood, but involves at least SyrM, another positive activator related to the LysR-NodD family. One of the phenotypically mutant Tn5 insertions used in genetic studies of NodD3-dependent *nod* regulation lacks NodD3 protein as determined by Western blots, but another expresses about 50–60% of the wild type level. The location of these Tn5 insertions substantially upstream of the open reading frame for NodD3 suggests importance of relatively distant regulatory sequences for *nodD3* expression. An insertion that did not cause a NodD3$^-$ phenotype is located in the extreme C-terminus of the protein coding region.

INTRODUCTION

Soil bacteria of the genus *Rhizobium* infect specific host plants and initiate the development of nitrogen-fixing root nodules. This symbiotic relationship is a complex interaction between each *Rhizobium* species and its limited group of host plants. *Rhizobium meliloti* nodulates alfalfa (*Medicago sativa*) and other related plants. Bacterial genes involved in the mechanism of nodulation are termed *nod* genes, and regulation of this complicated process depends on loci initially represented by *nodD* (1). NodD is required for expression of operons including *nodABC*, *nodFE* and *nodH* by a mechanism that involves an inducer from the plant host such as luteolin or other flavonoids (2, 3, 4, 5). NodD regulates expression of the inducible *nod* genes at the transcriptional level (6), and sequence analysis of NodD determined that it is a member of the LysR family of positive gene activator proteins (7). NodD binds DNA specifically at a ~50bp conserved region upstream from the *nod* genes (referred to as the '*nod* box') (8, 9, 10).

R. meliloti contains a multigene family of *nodD* loci, represented by *nodD1*, *nodD2* and *nodD3* (11). *Rhizobium leguminosarum* biovar *phaseoli* also has three copies of *nodD*, while other species of *Rhizobium* may have only one or two copies (12, 13). The species source or allele of *nodD* in the cell influences the response of *Rhizobium* to particular flavonoid inducers (14). In species having only one copy of *nodD*, *nodD* mutants are Nod$^-$, whereas in *R. meliloti* mutation of all three versions of *nodD* is required to result in a Nod$^-$ phenotype (11). The three *R. meliloti* NodDs are not functionally equivalent, as they differ not only in their response to plant inducers, but also in their regulatory properties. For example, luteolin is the strongest inducer of *nodABC* when *nodD1* is active, but *nodD2* responds to other plant signals and not to luteolin (15, 16). NodD3 is unique in that it can direct induction without plant inducers. The expression of *nodD3* is regulated by the closely linked symbiotic regulator, *syrM*, that is homologous to LysR-NodD proteins but appears not to be inducer dependent (16, 17). In this study, we determined the primary sequence of *nodD3* and examined its relationship to *nodD1*, *nodD2* and *nodD* genes of other *Rhizobium*. We verified the molecular weight of NodD3 by expression in *in vitro* transcription-translation assays, and we determined the site of several Tn5 insertional mutations giving a NodD3$^-$ phenotype to correlate genetic and physical properties with protein production.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions

Bacterial strains and plasmids used in this work are listed in Table 1. *E. coli* was grown in LB (21) supplemented with the appropriate antibiotics (ampicillin, 50 μg/ml; kanamycin, 25 μg/ml; tetracycline, 10 μg/ml) at 37°. Strains grown for preparation of single-stranded DNA were inoculated in 2XYT (21) medium supplemented with ampicillin and containing M13K07 (19) helper phage (4×10^7 phage/ml). After 30 min, kanamycin (70 μg/ml) was added to select for phage-infected

* To whom correspondence should be addressed

Table 1. Bacterial strains and plasmids used in this study

Strains	Relevant characteristics	Source/Reference
E. coli XL1-Blue	*recA*$^-$, *lac*$^-$, F′, Tc^r	Stratagene
R.meliloti RCR2011	Wild type	(18)
R. meliloti 1021	RCR2011 derivative, Str^r	(18)
JM88	1021 derivative, *nodD3*::Tn*5* #303, *nodC-lacZ*, Nm^r, Sp^r	(16)
JM204	1021 derivative, nodD3::Tn*5* #801, *nodC-lacZ*, Nm^r, Sp^r	(16)
Plasmids		
pUC118	Cloning vector	(19)
pUC119	Cloning vector	(19)
pBluescriptSK(+)	Cloning vector	Stratagene
pRmJT5	pLAFR1, Tc^r, *nodD3*$^+$	(5)
pRmJ30	pLAFR1, Tc^r, *nodD1*$^+$	(20)
pRmM144	5.3 kb *Xba*I-*Bam*HI fragment (*nodD3*) from pRmJT5 in pUC119	This study
pRmM147	2.1 kb *Cla*I fragment (*nodD3*) in pUC119	(16)
pBGR1	1.6 kb *Pst*I-*Bgl*II fragment(*nodD3*) from pRmM147 in pUC119, Ap^r	This study
pBGR2	1.6 kb *Pst*I-*Bgl*II fragment(*nodD3*) from pRmM147 in pBluescript,Ap^r	This study
pMY19	2.0 kb *Bgl*II fragment (*nodD3*) from pRm144 in pUC118	This study
pRmS303	pRmJT5::Tn*5*, Tc^r, Nm^r, *nodD3*$^-$	(5)
pRmS801	pRmJT5::Tn*5*, Tc^r, Nm^r, *nodD3*$^-$	(5)

cells. *R. meliloti* strains were grown in TY (22) medium with appropriate antibiotics (tetracycline, 10 μg/ml; neomycin, 50 μg/ml; spectinomycin, 50 μg/ml) at 30°.

Materials

Restriction enzymes and Exonuclease III were purchased from Promega Biotech and Bethesda Research Laboratories. T4 DNA ligase was obtained from New England Biolabs. Radiolabeled substrates were purchased from Amersham Corp.

Construction of *nodD3* subclones for sequencing

Restriction digestions, transformations, and alkaline lysis plasmid isolations were performed as described by Maniatis (23). Subclones of pRmM147 were prepared by *Pst*I (polylinker site) and *Bgl*II digestion and isolation of the insert fragment from an agarose gel by the freeze-squeeze method (24). The fragment was ligated into *Pst*I/*Bam*HI-digested pUC119 or pBluescript SK(+) to obtain subclones in opposite orientations for sequencing. pMY19 was obtained by ligation of a 2.0 kb *Bgl*II fragment from pRm144 into pUC118.

Subclones for determining locations of Tn*5* insertions in the *nodD3* region were constructed by utilizing a unique *Bam*HI site in Tn*5*. All subclones were obtained using isolated *Bam*HI-*Cla*I fragments ligated into *Bam*HI/*Cla*I-digested pBluescript. Subclones containing the fragment with the Tn*5* gene for neomycin resistance were selected on LB plates containing ampicillin and kanamycin. Subclones containing fragments with no selectable marker were selected on LB plates containing ampicillin and evaluated for possession of the appropriate fragment by restriction digestion.

Determination of Nucleotide Sequence

Nested deletions of pBGR1 and pBGR2 were prepared by ExonucleaseIII digestion as described by Henikoff (25). Single-stranded DNA was prepared as described previously (6), and both strands were sequenced by the dideoxy-chain termination method (26) using Sequenase (U.S. Biochemical Corp). Most reactions were performed using the universal primer. Two segments were sequenced using custom-made oligonucleotide primers as noted in Figure 1. Location of the Tn*5* insertions was determined by sequencing double-stranded DNA (27) using a primer (5′-GTT-CAGGACGCTACTTG-3′) complementary to DNA near the ends of Tn*5*.

Sequence Analysis

The amino acid sequence was compared to sequences in the National Biomedical Research Foundation (NBRF) database, Release 23, using FASTA (28). Nucleotide and amino acid sequences were compared using the GAP and BESTFIT programs from the University of Wisconsin GCG Sequence Analysis Software Package v6.1a (29). Multiple sequence alignments were obtained using TULLA version 2.1 (30). The TULLA program takes the supplied sequences and first performs all pairwise comparisons. The most related pair becomes the locked set against which the remaining sequences are compared to determine the next most similar sequence, which then becomes part of the locked set. This process continues until all of the sequences are aligned. The pairwise alignments are done using the Needleman-Wunsch method that compares the similarity score to a score obtained when an alignment against the randomized sequence is performed. Aligned sequences were analyzed for relatedness using the PROTPARS phylogeny program from PHYLIP version 3.02 (31), which infers an unrooted phylogeny by the parsimony method. This method is based on establishing a tree that requires the minimum number of mutational changes.

Expression of NodD3 protein

Single-stranded DNA from previously sequenced deletions was transformed into *E. coli* XL1-Blue in order to obtain double-stranded plasmid DNA. Plasmids purified by CsCl gradient centrifugation were incubated with extracts from *E. coli* HB101 and *R. meliloti* RCR2011 as previously described by Fisher *et al.* (6). Protein products were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and visualized by autoradiography.

Western Analysis of Tn*5* Mutants

A western blot was performed as described previously (32). The accompanying β-galactosidase assays were carried out according to Miller (21) with modifications as described by Mulligan and Long (16).

RESULTS AND DISCUSSION

Determination of *nodD3* Sequence and Coding Region

The 2.0 kb *Bgl*II fragment containing *nodD3* and surrounding sequences (Fig. 1) was subcloned and sequenced as described in Materials and Methods. An open reading frame (ORF) encoding a protein of 313 amino acids (expected molecular weight, 35.3 kDa) was evident from analysis of the sequence, as shown in Figure 2. The protein encoded by this ORF was found to be 77% identical to NodD1 (33) and 79% identical to NodD2 (34), whereas NodD1 and NodD2 are 87% identical to each other. Expression of the protein product encoded by this ORF was confirmed by *in vitro* transcription/translation assay using *R. meliloti* S-30 extracts (Fig. 3A), consistent with *in vivo* expression of *nodD3* clones in *E. coli* (16). The location of the *nodD3* coding region was confirmed using upstream deletions as shown in Figure 3B. The autoradiograph shows that a protein product migrating at the expected size of approximately 35 kDa is expressed when full-length pBGR2 or three increasingly smaller deletion clones are used. When a deletion into the putative coding region (Fig. 3A, lane 7) is utilized in the assay, no 35 kDa protein is obtained. In the controls containing no DNA and vector without insert, no 35 kDa product is generated.

Location of Tn5 Insertions in *nodD3* region

The *nodD3* locus was first identified by Tn5 insertions in pRmJT5 (5, 16). Three of these insertions (#303, #709, #801), mapping within the 2.0 kb *Bgl*II fragment of pRmJT5 containing *nodD3*, decreased the expression of a homogenotized *nodC-lacZ* fusion compared to the high constitutive expression caused by unmutagenized pRmJT5 (5). Restriction mapping indicated that these insertions were upstream of the *nodD3* coding sequence. Also, two downstream Tn5 insertions with no mutant phenotype mapped near the C-terminus of *nodD3*. To determine the exact locations of these mutations we sequenced each Tn5 junction with a Tn5-specific primer. Subclones were constructed as described in Materials and Methods.

As shown in Figure 2, none of the insertions which lead to a *nodD3* mutant phenotype are located within the coding sequence, suggesting that there may be important regulatory elements greater than 400 bp upstream of the translational start codon. In contrast, Tn5 #1005 interrupts the coding sequence of *nodD3* at the C-terminus, but does not exhibit the mutant phenotype, indicating that this region is not necessary for a functional NodD3 protein. This insertion changes the last residue (the codon immediately following the 9 bp repeat created by the transposition event) from arginine to proline, and extends the coding sequence an additional eight amino acids, creating the following C-terminal sequence: PDSYTQVAS. NodD proteins vary in length, and homology among NodD proteins is lost at the C-terminus (35, see also Fig. 5). The fact that an insertion at the end of NodD3 does not effect a mutant phenotype is another indication that this region is not critical for its activating function.

Unlike NodD3, most NodDs require an activating compound from the host plant to induce transcription of the inducible *nod* genes. One model of regulation by NodD proteins suggests that the N-terminal region of NodD contains the putative DNA binding domain (14, 35; see Fig. 5), and the C-terminus is the site of interaction with host-specific flavonoid compounds from the plant. However, some evidence from mutant studies indicates that this division of function is not straightforward and that there is some interaction between the N-terminus and C-terminus of these regulatory proteins (36, 37, 38). In addition, if the C-terminus does contain sequences important for flavonoid contact, NodD3 would be expected to show differences in sequence in this domain since it does not need an inducer to activate *nod* genes. However, the region where homology is lost between NodD3 and the other two NodD proteins from *R. meliloti* (noted by the asterisks in Figure 5) does not correspond to any point mutants recovered in genetic screens searching for altered NodD-flavonoid inducing activities (36, 37). Also, most of the residues in this segment are found in a NodD of one of the other *Rhizobium* species. This lack of a discernible pattern of critical residues from all of the NodD mutant studies indicates the complexity of this structure-function problem. Understanding these results will require further experiments such as flavonoid binding assays and detailed biochemical analysis of NodD structure to determine NodD-flavonoid contacts.

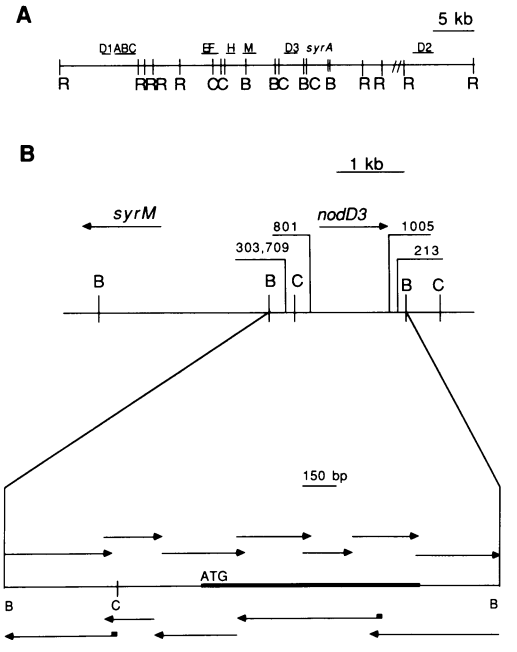

Figure 1. Location and sequencing strategy of the 2.0 kb *Bgl*II fragment containing *nodD3*. (A) Map of the *nod* gene region, adapted from Mulligan and Long (15). *Eco* RI sites are shown for the entire segment. The deletion indicated by slanted lines (//) is approximately 47 kb. *Cla*I and *Bgl*II sites are indicated only for the 15.5 kb *Eco* RI fragment containing *nodD3*. A methylated *Cla*I site internal to *nodD3* is not shown. The putative location of *syrA* is indicated by a dotted line. (B) Map of the *nodD3* region. Location of Tn5 insertions adjacent to *nodD3* are shown. Arrows represent sequence obtained from Exonuclease III-generated deletions of pBGR1 and pBGR2. The sequence of the upstream *Bgl*II-*Cla*I fragment was obtained using pMY19. The arrows with end boxes represent sequence obtained using specific oligonucleotide primers. The thick line indicates the coding region of *nodD3*. Abbreviations: R, *Eco* RI; C, *Cla*I; B, *Bgl*II.

Evaluation of Tn5 Mutants by Western Blot

Two of the Tn5 insertions that define the NodD3$^-$ phenotype were tested for the correlation of loss of function with loss of expression of the protein. Extracts were made from strains carrying the insertion in the genomic copy of *nodD3* as well as on a plasmid- borne copy of the gene; control strains contained pRmJT5 in each mutant background. A positive control for binding of the polyclonal anti-NodD1 antibody was provided by the wild-type strain containing pRmJ30 (NodD1). Strains JM88 and JM204 contain a *nodC-lacZ* fusion which allows us to compare loss of NodD3-dependent inducing activity in these

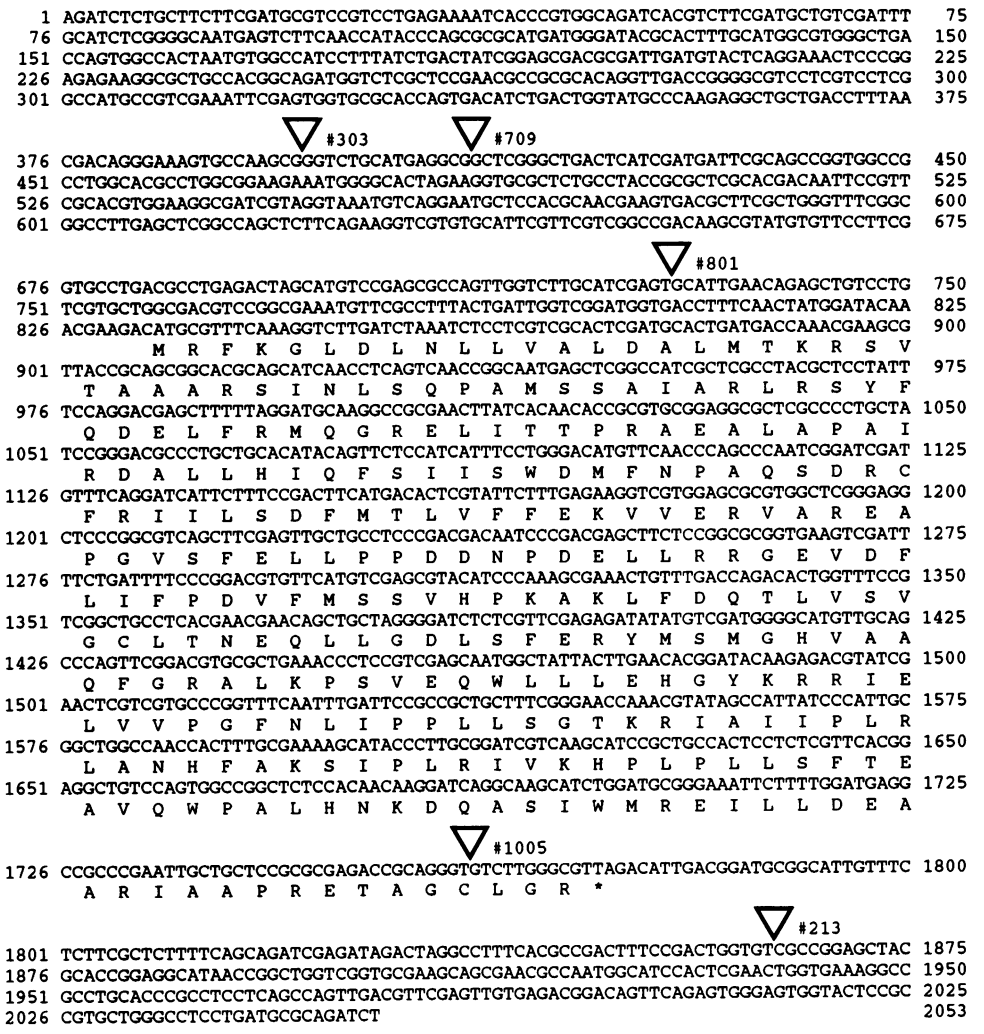

Figure 2. Nucleotide sequence and deduced amino acid sequence of *nodD3*. By convention, triangles indicate the 5′ end of the 9 bp repeat created by each Tn*5* insertion. Only one junction of Tn*5* #303 was determined. The *nodD3* sequence is available under EMBL accession number X53820.

Tn*5*-containing backgrounds (Table 2), JM88/pRmS303, containing Tn*5* #303, shows little NodD3 protein (Fig. 4, lane 4), corresponding to its near background level of β-galactosidase activity (Table 2). In contrast, JM204/pRmS801 shows a decrease in *nodC-lacZ* expression of approximately 40% from that produced by JM204/pRmJT5 (Table 2), indicating a leaky *nodD3* mutation. This phenotype is confirmed by the Western blot (Fig. 4, lane 2) in which NodD3 protein is observed in that strain, probably due to expression from a promoter within Tn*5*. Preliminary experiments on strains containing Tn*5* #709 (data not shown) produce a similar phenotype to Tn*5* #303 strains as predicted since these insertions are separated by only 15 base pairs. Previous reports postulated that insertion #303 is leaky because a $D1^{-}D2^{-}$ strain containing Tn*5* #303 has a delayed ability to nodulate alfalfa weakly (16); this nodulation ability does not appear to be a function of NodD3 protein as shown here by our expression data. However, the minor level of NodD3 production barely detectable by Western blot was enough to increase *nodC-lacZ* expression approximately 2−3 fold from background (Table 2) and may be sufficient for delayed nodulation; alternatively, expression of *nodD3* during the nodulation process may be different from that in our culture conditions (39).

Evolutionary Analysis of NodD Protein Sequences

Sequence comparison may provide clues to the ancestry of the NodD proteins as as well as information related to function. A search of the NBRF protein database revealed that, as expected, NodD3 shared homology to NodD sequences from other *Rhizobium* and *Bradyrhizobium* as well as to the previously described LysR family of regulatory proteins (7). No other significant homology was indicated.

The pattern of occurrence of *nodD* homologs across species is intriguing. Some strains contain one *nodD* gene while others contain two or even three versions of the gene (13), as in the case of *R. meliloti*. Since several *nodD* gene sequences from different species are now available in the database, we examined how the encoded proteins are related. Seven NodD sequences were aligned by the TULLA program (Fig. 5), and the alignment was subjected to the parsimony method of inferring unrooted phylogenetic trees, using PROTPARS from the PHYLIP version 3.02 package (31). A tree produced from only the three NodD sequences from *R. meliloti* (not shown) indicates that D1 and D2 are more related to each other than to D3. This result correlates with the relationship expected from simple percent identity. The phylogeny constructed using all the NodD sequences

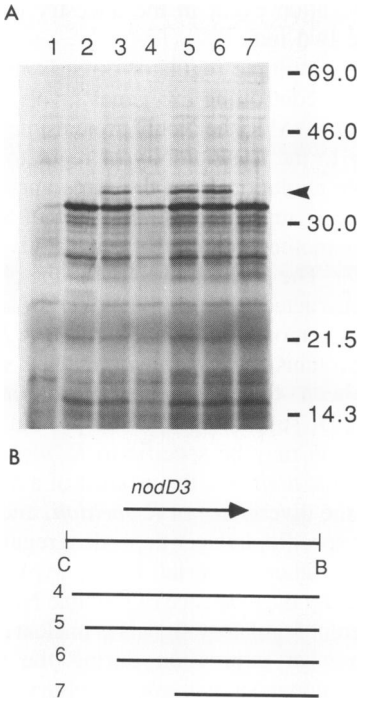

Figure 3. Expression of NodD3 *in vitro*. (A) SDS-polyacrylamide gel electrophoresis of coupled transcription-translation assays. Numbers on the right indicate Rainbow molecular weight markers (Amersham) in kilodaltons. The arrowhead shows the location of the *nodD3* product. The location of the *nodD3* open reading frame was confirmed by use of upstream deletions. Lanes 1: No added DNA. Lane 2: Vector alone. Lane 3: pBGR2. Lanes 4–7 are increasingly larger deletions of pBGR2. (B) Deletions of pBGR2 used in protein expression analysis. Lines indicate amount of *ClaI-BglII* fragment remaining in deletion clone. Numbers correspond to lanes in 3A.

Table 2. Effect of Tn*5* #303 and #801 on a *nodC-lacZ* fusion

Strain	β-galactosidase units[a]
JM88[b]	6.1
JM88/pRmS303	16.1
M88/pRmJT5	194.8
JM204[b]	16.8
JM204/pRmS801	143.2
JM204/pRmJT5	229.9

[a]Numbers represent the average of 2–4 assays.
[b]Data provided by J. Swanson.

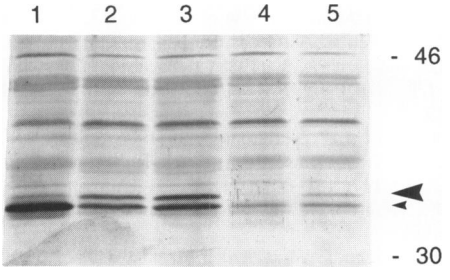

Figure 4. Western analysis of *nodD3*::Tn*5* mutants. Numbers indicate position of Rainbow molecular weight markers (kDa). Lanes: 1) *R. meliloti* 1021/pRmJ30, 2) JM204/pRmS801, 3) JM204/pRmJT5, 4) JM88/pRmS303, 5) JM88/pRmJT5. The large arrowhead indicates migration position of NodD3. The small arrowhead indicates position of NodD1. Note that lanes 1–3 are slightly overloaded.

```
                                    ^^^^^^^^^^^^^^^^^^^^
R.m.D1   MRFRGLDLNLLVALDALMTERKLTAAARRINLSQPAMSAAIARLRTYFGD
R.m.D2                                  VK
R.l.D       K                           S           S   D  R
R.m.D3  >   K                 K SV      S         S     S  Q
R.f.D2      K                 K SV      S
R.f.D1      K                           S             T    R
B.sp.D      K                   N       K                S  R

R.m.D1   ELFSMQGRELIPTPRAEALAPAVRDALLHIQLSVIAWDPLNPAQSDRRFR
R.m.D2                                                  I
R.l.D    D  I  R   V                E                   I   E
R.m.D3  >   R       T             I         F I S  MF        C
R.f.D2   D  T R           I                  F I S  MF  V  Q
R.f.D1      T N    V                E      H  I S   F         S
B.sp.D      T R    VL  G     GP    E         I SR AFD     S

R.m.D1   IILSDFMILVFFARIVERVAREAPGVSFELLPLDDDPHELLRRGDVDFLI
R.m.D2                E V                             Y          V
R.l.D           A     DK IL L                         E
R.m.D3  >       T     EKV                     P  N D       E
R.f.D2    R   VIM     E V K L       IG          TE  D     Y     V
R.f.D1          T M   E V V         A         FS E D
B.sp.D   V      TI    R   D I Q     A R       FS E S       E

R.m.D1   FPDVFMSSAHPKAKLFDEALVCVGCPTNKKLLGNISFETYMSMGHVAAQF
R.m.D2   L  L                  A              EQ   EL   K
R.l.D    L  L   G      R  E R                 EQ Q QL L Q          K
R.m.D3  >         V           QT  S   L       EQ   DL   R
R.f.D2   L EL A  D       L DT                   Q KRQL   N G     I  K
R.f.D1   L EM    HT  R       RF    S           Q  PPQL IDN V
B.sp.D   L EL           T     T                  Q SRPLT  K N      T
                                                                   *
R.m.D1   GREMKPSVEQWLLLEHGFNRRIELVVPGFTLIPRLLSGTNRIATLPLRLV
R.m.D2     AL    F             K  V                P  PH      II
R.l.D      GL           MQ     LK                N   P          I
R.m.D3  >  AL                   YK                N   P     K    II     A
R.f.D2     TL    I N            LK     V          S   P     D    M
R.f.D1     K-QR     E    R      LR   V VA         M   PF    D              A
B.sp.D      ALR NL E F          LK     V  Q   S    PI LD S   G M          A

           ** ***
R.m.D1   KYFEQTIPLRIVTSPLPPLF-FTEAIQWPALHNTDPGNIWLREILLQEAS
R.m.D2                 KH      W-     V         K       M
R.l.D     HY R       IEH    LVS-     V                  M    MI    L
R.m.D3  > NH AKS       KH     L S-     V         K QAS   M       D  A
R.f.D2    EH AK T    VAEL   A- PP AQ V    S    R QAS   M QV      L
R.f.D1    MH AKA      TEL Q IFPA      V     P SS  ASL M    F
B.sp.D    RH  KRM  Q  EP     - PT      V     F     AS   M R    E

R.m.D1   RIDPQSDTC-----------------------------
R.m.D2    SEF GE SLE----------------------------
R.l.D     MESEMECTS-----------------------------
R.m.D3  >  AAPRE AGCLGR---------------------------
R.f.D2   HMTTPP SVEYRP----------------------------
R.f.D1    VEF  E SAHALSSSQLPTCL-----------------
B.sp.D   NMASEHREPPTPQARLDSRPRRCKNILINHSTAWPASSF
```

Figure 5. Similarity of NodD3 to other NodD proteins. Alignment by TULLA (30). The complete sequence of *R. meliloti* NodD1 is shown. Amino acid changes in the other NodDs are indicated. NodD3 is marked by an arrowhead (>). Carets (^) indicate a putative helix- turn-helix DNA binding region (7). Asterisks (*) highlight a region of loss of homology between NodD3 and the other NodD proteins from *R. meliloti*. The C-terminal 9 amino acids of *R. leguminosarum* biovar *viciae* NodD have been corrected from the database version as shown in (42). Abbreviations: R. m., *Rhizobium meliloti*; R. l., *Rhizobium leguminosarum* bv. *viciae*; R. f., *Rhizobium fredii*; B. sp., *Bradyrhizobium* sp. ANU289.

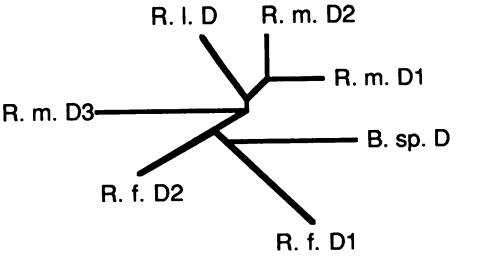

Figure 6. Schematic diagram of unrooted phylogenetic tree determined by PROTPARS (31). Approximate branch lengths were estimated using the minimum number of steps required between each branch point. Abbreviations as in Figure 5.

from the TULLA alignment is shown in Fig. 6. Since the parsimony method does not estimate branch lengths, these lengths were approximated using the minimum number of steps required for each branch as indicated by the program. The *R. meliloti* D1 and D2 sequences cluster, suggesting a relatively recent duplication event within this species. Parsimony analysis also reveals that NodD from *R. leguminosarum* bv. *viciae* is more similar to *R. meliloti* NodD1 and D2 than is NodD3.

That NodD1 and NodD2 from *R. meliloti* are functionally and physically related to the inducer-requiring NodD proteins of other fast-growing *Rhizobium* species indicates that these loci probably derived from an ancestral version of this regulatory gene. The *nodD1* locus is adjacent to the common *nod* genes [*nodABC*, genes which are structurally and functionally conserved across many *Rhizobium* species; see (1) for review] and is transcribed divergently from that operon, lending itself to consideration as the first regulator of the operon in *R. meliloti*. The fact that NodD3 is less similar to the other NodDs from its own species than it is to *R. leguminosarum* NodD could indicate that a lateral gene transfer has occurred, but this exchange seems unlikely considering the absence of inducer-independent NodD proteins in other species. The *nodD3* locus may therefore be the result of a duplication event of *nodD1* after the divergence of *R. meliloti*, and subsequent evolution away from regulation of activity by host-provided inducer.

Another study has directly evaluated *Rhizobium* relationships from NodD sequences in order to assess whether the bacteria have evolved in parallel with their respective host plants (40). Using a different approach to determine an unrooted phylogeny, the neighbor-joining method, Young and Johnston (40) report that *R. meliloti* and *R. leguminosarum* NodD proteins are highly related to each other and that the *Bradyrhizobium* sequences show considerable divergence from the NodD sequences of these two *Rhizobium* species. That previous study concludes a common ancestry for the *nodD* genes used in their comparison, with the addition of possible lateral gene transfer among *R. leguminosarum* biovars. Another phylogenetic study using comparison of 16S rRNA and nitrogenase genes in *Rhizobium* and other nitrogen-fixing bacteria also grouped together *R. meliloti* and two *R. leguminosarum* biovars (41). These relationships are confirmed in the tree produced from our parsimony analysis (Fig. 6) in which the NodD from the *Bradyrhizobium* strain appears to be relatively distant from the *R. meliloti* and *R. leguminosarum* proteins, probably reflecting the divergence of the fast-growing species from the slow-growing species.

Recently, another gene from *R. meliloti* involved in regulation of both *nod* genes and exopolysaccharide (*exo*) genes, called *syrM* (symbiotic regulator), has been shown to be 30% similar to NodD1 (primarily in the N-terminal putative DNA binding region) including conservative substitutions (17). In contrast, the NodD proteins of *R. meliloti* are almost 80% identical. A parsimony analysis showed that SyrM is distantly related to the three *R. meliloti* NodD proteins, but is more similar to NodD3 than to the other two NodDs. The *syrM* locus and *nodD3* also share several regions of DNA homology in their upstream regions (B. G. Rushing and M. J. Barnett, unpublished observations); the potential importance of these sequences for regulation is unknown although recent observations indicate that *syrM* and *nodD3* each activate the other's expression and that *syrM* also positively regulates its own expression (J. A. Swanson and S. R. Long, manuscript in preparation). The reciprocal control of *syrM* and *nodD3* and their close and divergent positions suggest a possible duplication event in the ancestry of the regulatory elements of the two loci.

SyrM acts as a multiple regulator of early nodulation events by also positively controlling *exo* genes involved in symbiosis. The homology of SyrM to the NodD proteins may just be a result of its similarity to the LysR family of regulatory proteins and may not indicate evolution from the same ancestral version of NodD. This idea is supported by the fact that SyrM is slightly more similar to another member of that family, NahR, than it is to NodD1 (17). The position of SyrM in the phylogenetic tree (not shown) constructed from all the NodD proteins used in this study indicates it is more similar to NodD from *Bradyrhizobium* than to the D proteins from *R. meliloti*. Only weak homology to *syrM* has been found in other *Rhizobium* species by hybridization (17), suggesting that the protein and its *nodD3* regulatory function may be specific to *R. meliloti*.

We postulate that *nodD3* is the product of a duplication event of *nodD1* after the divergence of *R. meliloti*, and that it evolved away from host-specific, inducer-dependent regulation of activity to control by another bacterial locus, *syrM*. The possible advantage gained by the existence of multiple NodD proteins that respond to different regulatory signals is unclear, but may reflect different patterns of expression during the time-course of nodulation. Additional studies are necessary to establish the mechanism and importance of the *nodD3-syrM* pathway of *nod* gene induction in *R. meliloti*.

ACKNOWLEDGEMENTS

We are indebted to many members of the lab for numerous discussions pertaining to this work. We are grateful to J.Swanson for strain construction and sharing data prior to publication; to J.Schwedock for advice on DNA sequencing; to M.Barnett for advice on DNA sequencing and transcription-translation experiments; to J.Ogawa for suggesting the Western blot experiment. We thank J.Mulligan for the pRmM144 construct, and we are grateful to J.Mountain for help in running the parsimony analysis program. We thank M.B., J.Mountain, and R.Fisher for critical reading of this manuscript and A.Bloom for its skillful preparation. This work was supported by Public Health Service grant RO1-GM30962 from the National Institutes of Health to S.R.L. B.G.R. was supported by a Training Grant in Cell and Molecular Biology from the National Institutes of Health to Stanford University. M.M.Y. was supported by a National Institutes of Health Postdoctoral Fellowship.

REFERENCES

1. Long, S.R. (1989) Annual Review of Genetics **23,** 483–506.
2. Faucher, C., Maillet, F., Vasse, J., Rosenberg, C., Van Brussel, A.A.N., Truchet, G., and Denarie, J. (1988) J. Bacteriol. **170,** 5489–5499.
3. Mulligan, J.T., and Long, S.R. (1985) Proc. Natl. Acad. Sci. USA **82,** 6609–6613.
4. Peters, N.K., Frost, J.W., and Long, S.R. (1986) Science **233,** 917–1008.
5. Swanson, J., Tu, J.K., Ogawa, J.M., Sanga, R., Fisher, R., and Long, S.R. (1987) Genetics **117,** 181–189.
6. Fisher, R.F., Swanson, J., Mulligan, J.T., and Long, S.R. (1987) Genetics **117,** 191–201.
7. Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. (1988) Proc. Natl. Acad. Sci. USA **85,** 6602–6606.
8. Fisher, R.F., and Long, S.R. (1989) J. Bacteriol.**171,** 5492–5502.
9. Fisher, R.F., Egelhoff, T.T., Mulligan, J.T., and Long, S. R. (1988) Genes Dev.**2,** 282–293.
10. Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A., and Kondorosi, A. (1986) Proc. Natl. Acad. Sci. USA **83,** 1757–1761.

11. Honma, M.A., and Ausubel, F.M. (1987) Proc. Natl. Acad. Sci. USA **84,** 8558–8562.
12. Davis, E.O., and Johnston, A.W.B. (1990) Mol. Microbiol. **4,** 921–932.
13. Rodriguez-Quinones, F., Banfalvi, Z., Murphy, P., and Kondorosi, A. (1987) Plant Mol. Biol. **8,** 61–75.
14. Spaink, H.P., Wijffelman, C.A., Pees, E., Okker, R.J.H., and Lugtenberg, B.J.J. (1987) Nature **328,** 337–340.
15. Györgypal, Z., Iyer, N., and Kondorosi, A. (1988) Mol. Gen. Genet. **212,** 85–92.
16. Mulligan, J.T., and Long, S.R. (1989) Genetics **122,** 7–18.
17. Barnett, M.J., and Long, S.R. (1990) J. Bacteriol. **172,** 3695–3700.
18. Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.E., and Ausubel, F.M. (1982) J. Bacteriol. **149,** 114–122.
19. Vieira, J., and Messing, J. (1987) Meth. Enzymol. **153,** 3–11.
20. Jacobs, T.W., Egelhoff, T.T., and Long, S.R. (1985) J. Bacteriol. **162,** 469–476.
21. Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor University Press, Cold Spring Harbor.
22. Beringer, J.E. (1974) J. Gen. Microbiol. **84,** 188–198.
23. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual.Cold Spring Harbor University Press, Cold Spring Harbor.
24. Tautz, D., and Renz, M. (1983) Analytical Biochem. **132,** 14–19.
25. Henikoff, S. (1984) Gene **28,** 351–359.
26. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74,** 5463–5467.
27. Wang, L.-M., Weber, D.K., Johnson, T., and Sakaguchi, A.Y. (1988) BioTechniques **6,** 839–843.
28. Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA **85,** 2444–2448.
29. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res.**12,** 387–395.
30. Subbiah, S., and Harrison, S.C. (1989) J. Mol. Biol. **209,** 539–548.
31. Felsenstein, J. (1985) Evolution **39,** 783–791.
32. Egelhoff, T.T., and Long, S.R. (1985) J. Bacteriol. **164,** 591–599.
33. Egelhoff, T.T., Fisher, R.F., Jacobs, T.W., Mulligan, J.T., and Long, S.R. (1985) DNA **4,** 241–248.
34. Honma, M.A., Asomaning, M., and Ausubel, F.M. (1990) J. Bacteriol. **172,** 901–911.
35. Horvath, B., Bachem, C.W.B., Schell, J., and Kondorosi, A. (1987) EMBO J. **6,** 841–848.
36. Burn, J., Rossen, L., and Johnston, A.W.B. (1987) Genes Dev. **1,** 456–464.
37. McIver, J., Djordjevic, M.A., Weinman, J.J., Bender, G.L., and Rolfe, B.G. (1989) Mol. Plant-Microbe Interactions **2,** 97–106.
38. Spaink, G.P., Wijffelman, C.A., Okker, R.J.H., and Lugtenberg, B.E.J. (1989) Plant Mol. Biol. **12,** 59–73.
39. Sharma, S.B., and Signer, E.R. (1990) Genes & Development **4,** 344–356.
40. Young, J.P.W., and Johnston, A.W.B. (1989) Trends Evol. Ecol. Sci (TREE) **4,** 341–349.
41. Hennecke, H., Kaluza, K., Thony, B., Fuhrmann, M., Ludwig, W., and Stackebrandt, E. (1985) Arch. Microbiol. **142,** 342–348.
42. Burn, J.E., Hamilton, W.D., Wootton, J.C., and Johnston, A.W.B. (1989) Mol. Microbiol. **3,** 1567–1577.