MPMI Vol. 9, No. 3, 1996, pp. 187–197. Publication no. M-1996-0126-01R. © 1996 The American Phytopathological Society

Novel and Complex Chromosomal Arrangement of *Rhizobium loti* Nodulation Genes

D. Barry Scott[1], Carolyn A. Young[1], Julie M. Collins-Emerson[1], Eric A. Terzaghi[1], Eva S. Rockman[1], Pauline E. Lewis[2], and Clive E. Pankhurst[2]

[1]Molecular Genetics Unit, Department of Microbiology and Genetics, Massey University, Palmerston North, New Zealand; [2]Biotechnology Division, Department of Scientific and Industrial Research, Palmerston North, New Zealand

Received 13 April 1995. Accepted 27 December 1995.

A mutational and structural analysis of *Rhizobium loti* nodulation genes in strains NZP2037 and NZP2213 was carried out. Unlike the case with other *Rhizobium* strains examined to date, *nodB* was found on an operon separate from *nodACIJ*. Sequence analysis of the *nodACIJ* and *nodB* operon regions confirm that *R. loti* common *nod* genes have a gene organization different from that of other *Rhizobium spp*. At least 4 copies of *nodD*-like sequences were identified in *R. loti*. The complete nucleotide sequence of one of these, *nodD3*, was determined. A new host-specific *nod* gene, *nolL*, was identified adjacent to *nodD3*. NolL shares homology with NodX and other *O*-acetyl transferases. Mutational analysis of the *nod* regions of strains NZP2037 and NZP2213 showed that *nodD3*, *nodI*, *nodJ*, and *nolL* were all essential for *R. loti* strains to effectively nodulate the extended host *Lotus pedunculatus*, but were not necessary for effective nodulation of the less restrictive host, *Lotus corniculatus*. Both *nodD3* and *nolL* were essential for *R. loti* strains to nodulate *Leucaena leucocephala*.

Additional keyword: nodulation gene rearrangements.

The development of nitrogen-fixing leguminous nodules is the result of two-way molecular signaling between the legume host and the *Rhizobium* symbiont. *Rhizobium* nodulation (*nod*) genes, with the exception of *nodD*, are not expressed in normal growth medium but are induced on addition of root or seed exudates to the medium (Innes et al. 1985; Mulligan and Long 1985; Rossen et al. 1985). The inducer molecules present in the exudates were identified as flavonoids and related compounds (Firmin et al. 1986; Peters et al. 1986; Redmond et al. 1986; Spaink et al. 1987; Kosslak et al. 1987). The chemical identity of the inducers varies for different *Rhizobium* spp. While isoflavones are potent inducers of *Bradyrhizobium nod* genes (Kosslak et al. 1987), they act as antagonists of *Rhizobium leguminosarum* bv. *viciae nod* gene induction (Firmin et al. 1986).

Activation of *nod* gene expression requires the presence of *nodD* regulatory gene(s) (Mulligan and Long 1985; Rossen et al. 1985; Spaink et al. 1987). The product of *nodD* binds to a conserved sequence (Hong et al. 1987; Fisher et al. 1988), the *nod* box (Rostas et al. 1986), located upstream of most *nod* operons. NodD functions as a host-specific determinant of nodulation, interacting in a strain-specific way with a spectrum of flavonoid or flavonoid-like molecules (Horvath et al. 1987; Spaink et al. 1987). Most strains of *Rhizobium* studied have multiple *nodD* genes, including *R. meliloti* (Honma and Ausubel 1987; Göttfert et al. 1986), *R. leguminosarum* bv. *phaseoli* (Davis and Johnston 1990), and *R. fredii* (Appelbaum et al. 1988). Györgypal et al. (1988) showed that the three *nodD* genes in *R. meliloti* contribute in a host-dependent manner to *nod* gene expression.

Expression of the *nod* genes results in the production of lipo-oligosaccharide Nod signal molecules that are relayed back to the plant host, resulting in nodule meristem initiation (Lerouge et al. 1990; Truchet et al. 1991). The Nod factors characterized are all β-1,4-linked tetramers or pentamers of D-glucosamine that are *N*-acylated on the terminal nonreducing residue and *N*-acetylated on the other residues, but they differ in the substituents linked to the chitin oligomer backbone (Lerouge et al. 1990; Spaink et al. 1991; Schultze et al. 1992; Price et al. 1992; Mergaert et al. 1993; Poupot et al. 1993). The elucidation of the structure of Nod factors opened the way to characterizing the biochemical functions of the *nod* gene products. The common *nodABC* genes are involved in the synthesis of the Nod factor oligosaccharide backbone (Bulawa and Wasco 1991; John et al. 1993; Röhrig et al. 1994; Atkinson et al. 1994). The host-specific *nod* genes encode products that decorate the backbone oligosaccharide (reviewed by Downie 1994). For example, NodPQ and NodH from *R. meliloti* specify the 6-*O*-sulfation of the reducing terminal glucosamine (Schwedock and Long, 1990; Roche et al. 1991) whereas NodL from *R. leguminosarum* bv. *viciae O*-acetylates the C-6 of the nonreducing terminal glucosamine (Bloemberg et al. 1994). NodF and NodE are involved in the

Corresponding author: D. Barry Scott, Molecular Genetics Unit, Department of Microbiology and Genetics, Massey University, Palmerston North, New Zealand; E-mail: d.b.scott@massey.ac.nz

Present address of Clive E. Pankhurst: CSIRO Division of Soils, Private Bag No. 2, Glen Osmond, South Australia 5064.

Present address of Eva S. Rockman: German-Israel Foundation for Scientific Research and Development, POB 7008, Jerusalem, Israel 91070.

Nucleotide sequence data will appear in the EMBL/GenBank/DDBJ nucleotide sequence data libraries under the accession numbers L06241, X65620, and U22899.

synthesis of polyunsaturated fatty acids that *N*-acylate the Nod factors (Spaink et al. 1991).

We previously identified a *nod* gene region in *R. loti* strain NZP2037 (Scott et al. 1985) that was subsequently shown to correspond to *nodACI* (Collins-Emerson et al. 1990). Given this novel organization of *nod* genes we decided to carry out a more detailed mutational and structural analysis of strain NZP2037, which forms effective nodules on a wide range of legumes, and a second strain, NZP2213, which forms effective nodules on a more limited legume host range (Pankhurst et al. 1979, 1987). In this report we show that the structural organization of *R. loti* common *nod* genes is significantly different from that of other *Rhizobium* spp. (Young and Johnston 1989) with *nodB* on an operon separate from *nodACIJ*. In

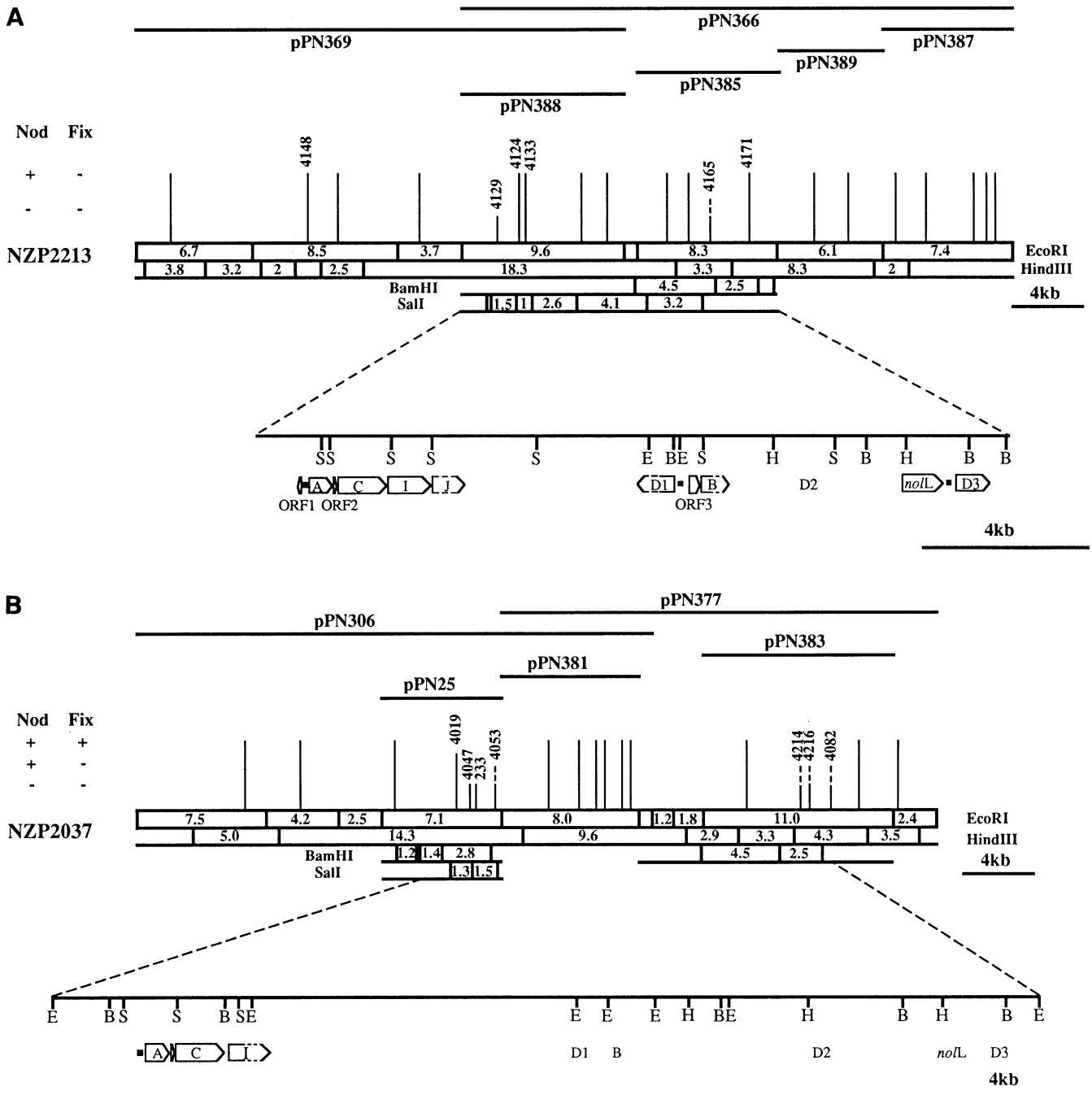

Fig. 1. Physical and genetic maps of the nodulation (*nod*) gene region of *Rhizobium loti* strains NZP2213 and NZP2037. Restriction enzyme sites for *Eco*RI (E), *Hin*dIII (H), *Sal*I (S), and *Bam*HI (B) are shown for the *nod* regions of **(A)** NZP2213 and **(B)** NZP2037. The organization of the *nod* genes as determined by either DNA sequence (boxed arrows) or hybridization (unboxed) is also shown. Boxes with continuous lines represent completely sequenced genes whereas discontinuous lines represent partially sequenced genes. Filled boxes represent *nod* boxes. The sites of various Tn*5* insertions, and in some cases the allele number, are shown as vertical bars. The nodulation (Nod) and fixation (Fix) phenotype (scored as + or -) of the corresponding *R. loti* mutants inoculated on *Lotus pedunculatus* seedlings is indicated by the height of the bar.

addition we report on the identification and sequence analysis of a new host specificity gene from *R. loti*, designated *nolL*, that shares sequence similarity to *nodX* (Davis et al. 1988).

RESULTS

Isolation of *nod* cosmids from *R. loti* NZP2213 and NZP2037.

Through the use of a previously isolated Nod$^-$ mutant (strain PN233) of NZP2037 as recipient and an *Escherichia coli*/pLAFR1 gene library of NZP2213 as donor, the corresponding *nod* region of strain NZP2213 was isolated by in planta complementation (Long et al. 1982). When *Eco*RI digests of a representative set of the complementing cosmids were analyzed on an agarose gel, two patterns of overlapping fragments were observed, corresponding to the *Eco*RI fragments of pPN369 and pPN366 (Fig. 1A). An *Eco*RI and *Hin*dIII restriction enzyme map of this NZP2213 chromosomal region was constructed by comparing single and double digests of pPN369 and pPN366 and by probing Southern blots of these digests with appropriately selected *Hin*dIII or *Eco*RI fragments (Fig. 1A).

In parallel, the *nod* region of the wider host range strain NZP2037 was also analyzed. Cosmid pPN306 was previously isolated by in planta complementation of PN233 (Scott et al. 1985) and a physical map constructed (Fig. 1B). To extend this map over the same region as that for NZP2213 an additional cosmid, pPN377, was isolated from a gene library of NZ2037 by using pPN381 as a probe. A *Hin*dIII and *Eco*RI restriction enzyme map to pPN377 was constructed as described above for the NZP2213 cosmids. A complete physical map for the extended *nod* region of strain NZP2037 is shown in Figure 1B.

Tn*5* mutagenesis of NZP2037 and NZP2213 *nod* regions.

To identify the location of *nod* genes, other than NZP2037 *nodC* previously identified (Collins-Emerson et al. 1990), Tn*5* mutagenesis of *E. coli* strains containing pPN306, pPN377, pPN369, and pPN366 was carried out as described in Materials and Methods. The positions of the Tn*5* insertions are shown in Figure 1. Each unique Tn*5* cosmid insertion was then exchanged into the genome of NZP2037 or NZP2213 by homologous recombination and the symbiotic phenotype determined. Of the 37 NZP2037 Tn*5*-containing mutants tested, six (corresponding to insertions 4019, 4047, 4053, 4214, 4216, and 4082) showed an altered symbiotic phenotype on *Lotus pedunculatus* (Fig. 1B and Table 1). Mutations 4019, 4047, and 4053 all mapped close to the previously isolated *nodC233* mutation (Collins-Emerson et al. 1990) and sequence analysis of Tn*5* junctions showed that they correspond to insertions upstream of the start of *nodA* (224 bp upstream), within *nodC* (28 bp downstream) and within *nodI* (35 bp downstream), respectively (Table 1). PN4019 was Fix$^+$ but produced only 30 to 50% the dry weight of plants inoculated with wild type. PN4047 was defective in hair curling (Hac$^-$) and nodulation (Nod$^-$) whereas PN4053 was Hac$^+$ but had a 1 to 2 week delay in nodulation (Nod^d). All three mutations were fully complemented for nodulation of *Lotus pedunculatus* by introduction of either pPN306 (NZP2037) or pPN369 (NZP2213). The other three mutant strains, PN4214, PN4216 and PN4082, all had a delayed nodulation (Nod^d) phenotype and the Tn*5*s mapped to a 4.3-kb *Hin*dIII fragment, about 20 kb distant from *nodC233* (Fig. 1B). Sequence analysis of the Tn*5* junctions showed that they correspond to insertions in *nolL* (a new nodulation gene, see below), *nolL*, and *nodD3*, respectively (Table 1). These three mutations were complemented by introduction of either pPN377 (NZP2037) or pPN366 (NZP2213). When tested on the indeterminate nodulating host *Leucaena leucocephala*, PN4047, PN4214, PN4216, and PN4082 were all Nod$^-$, whereas PN4019 and PN4053 were Fix$^+$ (Table 1). On *Lotus corniculatus*, a determinate host like *L. pedunculatus*, only PN4047 showed an altered symbiotic phenotype (Nod$^-$) with the other mutants all being fully effective (Table 1).

For strain NZP2213, 20 unique Tn*5*-containing mutants were tested on plants, and of these only two, PN4129 and PN4165, showed an altered nodulation phenotype on *Lotus pedunculatus* (Table 1). Strain PN4129 was Hac$^-$ Nod$^-$ and corresponds to a Tn*5* insertion in *nodC* (Table 1), whereas PN4165 showed a variable response with some plants being Nod$^-$ and others Nod^d. As the wild-type strain, NZP2213, forms tumor-like Fix$^-$ nodules on *L. pedunculatus* (Pankhurst et al. 1979), this host was not suitable for scoring the Fix phenotype of the mutants, so instead *L. corniculatus* was used. By means of this host, a third mutant, strain PN4148, was identified that formed abundant small white spherical Fix$^-$ nodules. Strain PN4129 was Nod$^-$ on both *Lotus corniculatus* and *Leucaena leucocephala* (Table 1). All three mutations were complemented with cosmids from either NZP2037 or NZP2213; strain PN4129 with pPN306 (NZP2037), pPN369 (NZP2213), and pPN366 (NZP2213); strain PN4165 with either pPN377 (NZP2037) or pPN366 (NZP2213); and PN4148 with either pPN306 (NZP2037) or pPN369 (NZP2213).

Complementation of *R. leguminosarum* bv. *trifolii nod* mutations.

To better define the *R. loti* symbiotic loci identified by transposon Tn*5* mutagenesis, cosmids containing wild-type *nod* alleles were transferred to several *R. leguminosarum* bv. *trifolii* Nod$^-$ mutants and the transconjugants tested for

Table 1. Symbiotic phenotype of *Rhizobium loti* mutants

Strain	Plant host		
	Lotus pedunculatus	*Leucaena leucocephala*	*Lotus corniculatus*
NZP2037 derivatives			
NZP2037	Fix$^+$	Fix$^+$	Fix$^+$
PN4019	Fix^{+a}	Fix$^+$	Fix$^+$
PN4047 (*nodC*)[b]	Nod$^-$	Nod$^-$	Nod$^-$
PN233 (*nodC*)[b]	Nod$^-$	Nod$^-$	Nod$^-$
PN4053 (*nodI*)[b]	Nod^{d}Fix$^-$[c]	Fix+	Fix$^+$
PN4214 (*nolL*)[b]	Nod^{d}Fix$^-$[c]	Nod$^-$	Fix$^+$
PN4216 (*nolL*)[b]	Nod^{d}Fix$^-$[c]	Nod$^-$	Fix$^+$
PN4082 (*nodD3*)[b]	Nod^{d}Fix$^-$[c]	Nod$^-$	Fix$^+$
NZP2213 derivatives			
NZP2213	Nod$^+$Fix$^-$	Nod$^+$Fix$^-$	Fix$^+$
PN4148	Nod$^+$Fix$^-$	Nod$^+$Fix$^-$	Nod$^+$Fix$^-$
PN4129 (*nodC*)	Nod$^-$	Nod$^-$	Nod$^-$
PN4165	Nod$^-$/Nod^d	Nod$^-$	Fix$^+$

[a] Fixation 50% of wild type.
[b] Gene identified by DNA sequence analysis of Tn*5* junctions.
[c] Delay in nodulation (Nod^d) of 1 to 2 weeks.

Table 2. Complementation of *Rhizobium leguminosarum* bv. *trifolii* (red clover nodule phenotype) nodulation mutations

Plasmid	*R. leguminosarum* bv. *trifolii* mutant			
	ANU277 (*nodC*)[a]	ANU249 (*nodB*)	ANU252 (*nodA*)	ANU851 (*nodD*)
NZP2037 (*R. loti*)				
pPN306	++[b]	–	–	–
pPN377	–	–	–	±
pPN25	–	–	–	–
pPN381	–	–	–	–
pPN383	–	–	–	±[c]
NZP2213 (*R. loti*)				
pPN369	++	–	–	–
pPN366	++	+[d]	–	±
pPN388	++	–	–	–
pPN385	–	–	–	±
pPN389	–	–	–	–
pPN387	–[e]	–	–	–
PN100 (*R. leguminosarum* bv. *trifolii*)				
pPN26	++	++	++	++

[a] See Djordjevic et al. (1985).
[b] Uniform complementation; Nod$^+$Fix$^+$ nodules.
[c] Variable complementation; some plants Nod$^-$ but others Fix$^-$ or Fix$^+$.
[d] Uniform complementation; Nod$^+$Fix$^+$ nodules but fixation 20% of wild type.
[e] No complementation; Nod$^-$.

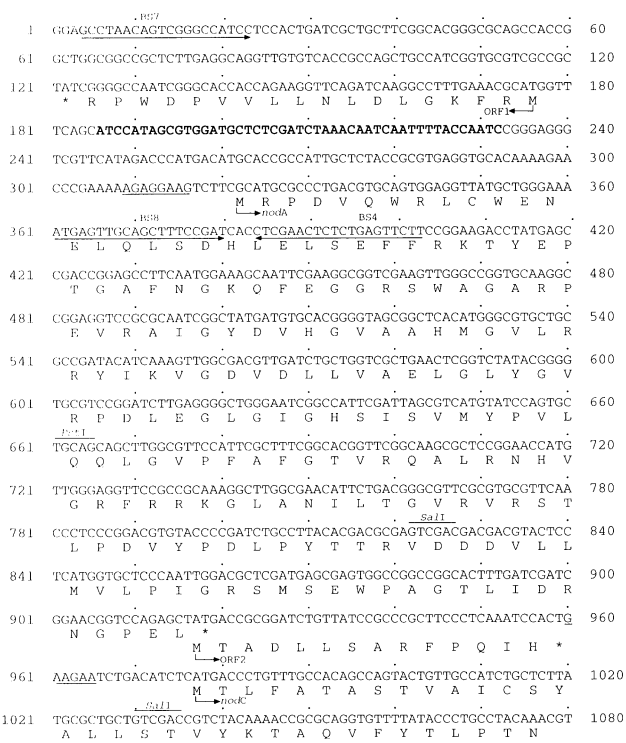

Fig. 2. DNA sequence and deduced amino acid sequence of *nodA* from strain NZP2213. The 1,080-bp region shown includes the following: the 5′ region (divergently read) of an open reading frame (ORF), ORF1, that corresponds to the N terminus of *nodD1*; an intergenic region containing a *nod* box consensus sequence (bold); the structural gene of *nodA*; an ORF, ORF2, corresponding to a C-terminal 45-bp duplication of *nodB*, (positions 917 to 961); a short intergenic region; and the N terminus of *nodC* (positions 977 to 1080). Ribosomal binding sites and primer sites for DNA sequencing are underlined. This sequence data will appear in the EMBL/GenBank/DDBJ nucleotide sequence data libraries under the accession number L06241.

nodulation on red clover. The results of these experiments are shown in Table 2.

Complementation of *R. leguminosarum* bv. *trifolii nodB* and *nodC* mutations was observed but not that of *nodA*. The fact that only pPN366 was able to complement the *nodB* mutation, whereas pPN306, pPN369, pPN366, and pPN388 all complemented *nodC*, indicated that *nodB* was separated from *nodC*. This was subsequently confirmed by DNA sequencing (see next section). In the case of the *nodD* transconjugants, a range of symbiotic phenotypes was observed, indicating that cross complementation of this mutation is inefficient.

Identification and sequence analysis of *R. loti nodACIJ*.

Given that the NZP2213 cosmid, pPN388 (Fig. 1A), complemented the *R. leguminosarum* bv. *trifolii nodC* mutation (Table 2), it was clear that at least *nodC* was located on this fragment. Single-stranded sequence analysis of the end regions of the *Eco*RI/*Sal*I and *Sal*I fragments of pPN388 (Fig. 1A) identified open reading frames (ORFs) similar to *nodA*, *C*, *I*, and *J* from other rhizobia (Jacobs et al. 1985; Egelhoff et al. 1985; Evans and Downie 1986).

The location of these ORFs and their presumed direction of transcription is shown in Figure 1. As *R. loti* DNA spanning this region (i.e., pPN366) was unable to complement a *R. leguminosarum nodA* mutation (see above) the DNA sequence of the NZP2213 *nodA* was determined to confirm the presence of a complete gene (Fig. 2). The NZP2213 *nodA* gene consists of a coding sequence of 594 bp starting with an ATG at position 324 and ending with a TGA (stop) codon at position 920 (Fig. 2). The deduced amino acid sequence of NodA contained 198 residues with an unmodified molecular mass of 22.2 kDa. Upstream of *nodA* is a 47-bp sequence (residues 186 to 232; Fig. 2) similar to the *Rhizobium nod* box sequence (Rostas et al. 1986). Farther upstream, but divergently "read," is a short ORF of 57 bp (176 to 120), designated ORF1. The deduced amino acid sequence of ORF1 shares 66% identity with the first 18 amino acids of *R. meliloti* NodD1 (Egelhoff et al. 1985). There is no further similarity either at the DNA or amino acid sequence level beyond the stop codon (Fig. 2).

Immediately downstream of *nodA* is a short ORF of 45 bp (917 to 961), designated ORF2. The deduced amino acid sequence of ORF2 had no significant similarity with any other known polypeptide sequences. However, at the DNA level there is 59% identity between ORF2 and the C terminus of *R. meliloti nodB* (Egelhoff et al. 1985). The lack of similarity at the polypeptide level is a consequence of a shift in the translational reading frame (Fig. 2). Downstream of ORF2 is a short intergenic region, then an ORF corresponding to the N terminus of NZP2213 NodC (residues 977 to 1080; Fig. 2). Further downstream is *nodI*, the sequence of which was previously reported (Young et al. 1990).

By means of a similar strategy to analyze the corresponding *nod* region of strain NZP2037, the gene order in this strain was shown to be identical to NZP2213, i.e., *nodA*-ORF-*nodCI* (Fig. 1B). The complete double-stranded sequence of the NZP2037 *nodC* has been published (Collins-Emerson et al. 1990).

R. loti nodB.

To identify the location of *R. loti nodB*, a Southern blot of *Eco*RI/*Hin*dIII double digests of pPN377 and pPN366 was

probed with a 1,220-bp *Eco*RI/*Kpn*I fragment containing *nodB* from *R. leguminosarum* bv. *trifolii* ANU843 (Schofield and Watson 1986). This probe hybridized to a 1.2-kb *Eco*RI fragment from NZP2037 (Fig. 1B) and a 2.3-kb *Eco*RI/*Hin*dIII fragment from NZP2213 (Fig. 1A). Sequence analysis of the *Eco*RI end of the 2.3-kb fragment from NZP2213 (Fig. 3) revealed the presence of an ORF similar to *nodB* from *R. meliloti* (Egelhoff et al. 1985). Immediately upstream of the putative ATG start codon for *nodB* (505 to 507; Fig. 3) is an ORF of 240 bp, designated ORF3, that is preceded by a good ribosome binding site (256 to 262; Fig. 3) and a *nod* box (residues 168 to 214; Fig. 3). The C-terminal 48 bp of ORF3 (461 to 508; Fig. 3) share 70 and 77% identity with the C-terminal DNA sequences of *R. loti* (Fig. 2) and *R. meliloti* (Egelhoff et al. 1985) *nodA*s, respectively. Farther upstream of the *nod* box is a sequence beginning at residue 148, but divergently "read" to *nodB*, that shares considerable similarity (77% for the 148 bp shown in Figure 3) to *R. meliloti nodD1* (Egelhoff et al. 1985). This *nodD*-like sequence, designated *nodD1*, lacks the N-terminal 54 nucleotides found in *R. meliloti nodD1*. The absence of the N terminus, together with the presence of several frame-shift mutations within the sequence, would suggest that this gene is inactive. Hybridization experiments suggest that there are at least four *nodD*-like sequences in the *R. loti* genome; three within the genomic region shown in Figure 1 and one outside (results not shown). The isolation of a symbiotically defective mutation (*nod*4082) in *nodD3* of NZP2037 (Table 1) confirmed that at least this copy was functional in *R. loti*.

***R. loti* host specificity genes *nolL* and *nodD3*.**

The isolation of Tn*5* insertions 4214, 4216, and 4082 in strain NZP2037 (Fig. 1B) identified a third symbiotic locus in *R. loti*. Sequence analysis of the Tn*5* junctions of these insertions demonstrated that 4214 and 4216 were located in a new symbiotic gene, designated *nolL*, whereas 4082 corresponded to an insertion in a *nodD* gene, designated here as *nodD3*. Mutations in both genes gave rise to a Nod$^-$ phenotype on the indeterminate host *L. leucocephala* and Fix$^-$ nodules on *L. pedunculatus*, indicating that they are host-specificity genes (Table 1). Complementation analysis (described above) indicated that both genes are functional in strain NZP2213 as well as NZP2037. We therefore decided to determine the DNA sequence of *nolL* and *nodD*3 from the *R. loti* type strain, NZP2213.

The DNA sequence of *nolL* and *nodD*3, spanning a region of 2,726 nucleotides, is shown in Figure 4. The *nolL* gene consists of a coding sequence of 1,119 bp starting with an ATG at position 191 and ending with a TAG (stop) codon at position 1312 (Fig. 4). The deduced amino acid sequence of NolL contained 373 residues with an unmodified molecular mass of 41.9 kDa. A BLAST search (Altschul et al. 1990) of the GenBank and EMBL data bases revealed that NolL shared significant similarity (Fig. 5) with an *O*-acetyl transferase from *Xanthomonas campestris* (Y. S. Lin et al., unpublished results), a deduced protein from *Salmonella enterica* (Wang et al. 1992) and an *O*-antigen acetylase from *Shigella flexneri* bacteriophage Sf6 (Clark et al. 1991), with BLAST scores of 64, 62, and 48, respectively. Clark et al. (1991) previously showed similarity between the Sf6 acetylase and *R. leguminosarum* bv. *viciae* NodX (Davis et al. 1988). An alignment of the amino acid sequences of these proteins is shown in Figure 5.

The *nodD3* gene consists of a coding sequence of 903 bp starting with an ATG at position 1547 and ending with a TGA (stop) codon at position 2452 (Fig. 4). The deduced amino acid sequence of NodD3 contained 301 residues with an unmodified molecular mass of 34.3 kDa. *Rhizobium loti* NodD3 shares 69% identity (81% similarity) with *R. meliloti* NodD1 (Egelhoff et al. 1985). Upstream of the proposed translational start site of *nodD3* is a poorly conserved *nod* box (residues 1491 to 1537).

DISCUSSION

The common *nod* genes in *R. meliloti* and *R. leguminosarum* are organized in a single operon as *nodABCIJ* with an adjacent *nodD* that is divergently transcribed (reviewed in Long 1989). In contrast, the *nodB* gene in *R. loti* strains NZP2037 and NZP2213 is separate from *nodACIJ*. In strain NZP2213 a remnant, 54 bp, of the N terminus of a *nodD* gene was found upstream of the *nodACIJ* operon. What appears to be the remaining C terminus of this *nodD* gene, *nodD1*, was located upstream of the *nodB* operon (Fig. 3). The presence of a C-terminal duplication of *nodA* (of about 45 to 50 bp) immediately upstream of *nodB*, and a C-terminal duplication of *nodB* (of about 45 to 50 bp) immediately downstream of *nodA*, suggests that *nodB* was once joined to *nodA* in *R. loti*. Assuming *nodABCIJ* was the ancestral operon organization (Young and Johnston 1989), then *R. loti* strain NZP2213 has undergone at least two rearrangements to generate a split *nodD1* gene and duplicated C termini of *nodA* and *nodB*. The *nod* genes of *R. loti* strain NZP2037 appear to have undergone a similar rearrangement (this work and Collins-Emerson et al. 1990).

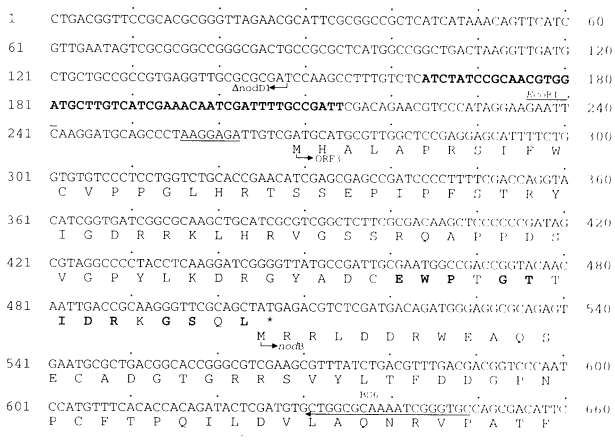

Fig. 3. DNA sequence and deduced amino acid sequence of strain NZP2213 *nodD1-nodB* intergenic region. The 660-bp region shown includes the following: part of an N-terminal deleted (54-bp) *nodD1* sequence (divergently read from positions 148 to 1); a *nod* box consensus sequence (bold); a 240-bp open reading frame (ORF), ORF3, that contains a 48-bp (positions 461 to 508) sequence similar (identical amino acids are shown in bold) to the C terminus of *R. meliloti nodA*; and the 5′-end of *nodB* (505 to 660). A putative ribosome binding site (256 to 262) and primer sites for DNA sequencing are underlined. The primer site for BS5 lies immediately upstream (–20 to –1) of the sequence shown. This sequence will appear in the EMBL/GenBank/DDBJ nucleotide sequence data libraries under the accession number X65620.

A complex rearrangement of *nod* genes has also been observed for the Type I group of *R. leguminosarum* bv. *phaseoli* (Vázquez et al. 1991), recently reclassified as *R. etli* (Segovia et al. 1993). In this case *nodA* is separated from *nodBC* genes. While there are regions of similarity between the *nodA* region and the ORF1-ORF2-*nodB-nodC* operon in *R. etli* (Vázquez et al. 1991), there are no obvious direct repeat sequences in the vicinity of the rearranged genes, as is found in *R. loti.* The 45 to 50 bp C-terminal duplications of *nodA* and *nodB* in *R. loti* provide target sites for further recombination.

Insertional mutagenesis and complementation experiments confirmed that the NZP2213 *nodB* and *nodC* genes identified here are functional. Complementation of *R. leguminosarum* bv. *trifolii nodC* with a *R. loti* cosmid (pPN388) that lacks a functional *nodD* would indicate that the *R. leguminosarum* NodD can substitute for *R. loti* NodD in activating the *R. loti nodACIJ* operon. In contrast, *R. loti nodD3* only weakly complemented a *R. leguminosarum nodD* mutation. The single *R. loti* cosmid, pPN366, known to contain both *nodACIJ* and *nodB* operons failed to complement the *nodA* mutation of *R. leguminosarum.* One possible explanation for this lack of complementation is that the Nod factor produced by ANU252 (*nodA*::Tn*5*) is a poor substrate for the *R. loti* NodA. In support of this hypothesis the activity of the *N*-acyltransferase (NodA) from *R. meliloti* has been shown to be sensitive to oligosaccharide chain length (Atkinson et al. 1994).

All three operons examined here, *nodACIJ, nodB,* and *nodD3,* contained classical (Rostas et al. 1986) *nod* box consensus sequences in their promoter regions. In the case of the *nod* box upstream of *nodD3,* however, only about half of the classical consensus is conserved. Recent work by Goethals et al. (1992) suggests that the functional unit of a *nod* box is a 15-bp inverted repeat sequence of ATC-N9-GAT, which like all members of the LysR-type of regulated promoters (Henikoff et al. 1988) is the core DNA binding motif. Two copies of this core sequence are present within the 47-bp *nod* box sequence identified upstream of the *nodACIJ* and *nodB* promoters of *R. loti.*

As in the case of *R. meliloti* (Honma and Ausubel 1987; Göttfert et al. 1986), there are multiple copies of *nodD*-like sequences in *R. loti.* As discussed above, *nodD1* is presumably inactive due to the absence of the N-terminal 54 bp and the presence of a series of frame-shift mutations. No information is yet available on *nodD2.* The third copy of *R. loti nodD* is host specific, being essential for *R. loti* to form effective symbioses with *L. pedunculatus* and *L. leucocephala* but not

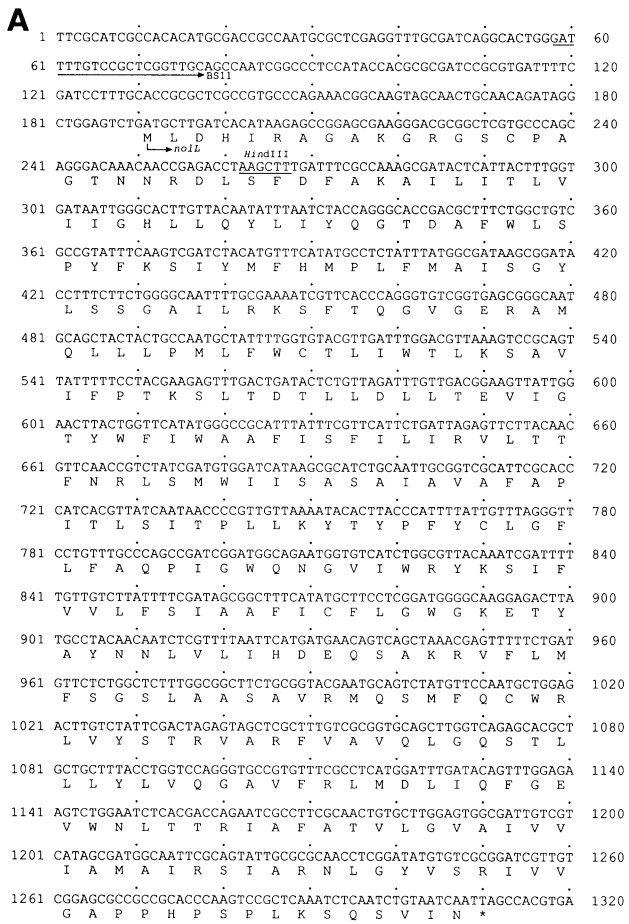

Fig. 4. DNA sequence and deduced amino acid sequence of *nolL* and *nodD3* from strain NZP2213. The 2,726-bp sequence shown includes the following: the 5′ region of *nolL*; the coding region of *nolL* (191 to 1309); the *nolL-nodD3* intergenic region including a *nod* box sequence (bold); the coding region of *nodD3* (1547 to 2449). The sequence ends at the *Eco*RI site shown in Figure 1A. Ribosomal binding sites for *nolL* (174 to 180) and *nodD3* (1534 to 1540) and the primer site for BS11 are underlined. This sequence will appear in the EMBL/GenBank/DDBJ nucleotide sequence data libraries under the accession number U22899.

with *L. corniculatus*. By comparison with what has been found for the function of multiple copies of *nodD* in *R. meliloti* (Györgypal et al. 1988; Györgypal et al. 1991), it is likely that the different NodDs in *R. loti* have different sensitivities to specific host flavonoids and consequentially activate *nod* gene expression in a host-dependent manner.

As is the case for *nodD3*, the requirement for a functional *nodI* gene product in *R. loti* is host dependent. NodI is required for forming an effective symbiosis on *L. pedunculatus* but is not essential for effective nodulation of *L. corniculatus*. In *R. leguminosarum*, mutations in *nodI* and *J* result in delayed nodulation of peas (Evans and Downie 1986) and deletion of *nodI* and *J* in *Bradyrhizobium japonicum* (Göttfert et al. 1989) results in strains with a Nod$^+$ phenotype when inoculated onto soybeans. Delayed nodulation was also observed for a *R. loti nodI* mutant when tested on *L. pedunculatus*, but in addition this mutant was also Fix$^-$ on this host. This phenotype may of course be an indirect consequence of *nodI* gene inactivation, as any delay in nodulation of *L. pedunculatus* results in a rapid accumulation of condensed tannins in this particular host, which in turn blocks infection and establishment of a nitrogen-fixing symbiosis (Pankhurst et al. 1979). Taken together, these results would suggest that *nodI* and *J* are important in improving the efficiency of nodulation on certain hosts and this is consistent with the recent demonstration that NodI and NodJ are involved in the secretion of Nod factors (McKay and Djordjevic 1993).

Adjacent to *nodD3* a new host-specific nodulation gene was identified, *nolL*. This gene is essential for *R. loti* NZP2037 to nodulate both *L. pedunculatus* and *L. leucocephala* but is not required for nodulation of *L. corniculatus*. Based on sequence comparisons, and by analogy with the recently demonstrated biochemical function of NodX (Firmin et al. 1993), we would predict that *nolL* encodes an acetyl transferase that *O*-acetylates the C-4 position of the fucose residue located on the reducing terminal glucosamine of the recently determined *R. loti* Nod factor (López-Lara et al. 1995). Alternatively, like NodZ in *Bradyrhizobium japonicum* (Stacey et al. 1994), NolL may be involved in adding the entire 4-acetylfucose.

In conclusion, the information presented here on the basic organization of both common and host-specific *nod* genes in *R. loti* will be very useful for comparative studies with other rhizobia, but more importantly should enhance the development of the model *R. loti–Lotus japonicus* symbiotic system (Handberg and Stougaard 1992).

MATERIALS AND METHODS

Bacterial strains and plasmids.

These are described in Table 3.

Growth of bacteria.

Rhizobium loti cultures were grown at 28°C in either TY medium (Beringer 1974) or S20 defined medium (Chua et al. 1985) supplemented where necessary with neomycin (Neo; 200 μg/ml for NZP2213 and 400 μg/ml for NZP2037), tetracycline (Tet; 2 μg/ml), gentamycin (Gen; 50 μg/ml) or streptomycin (Str, 200 μg/ml). *Rhizobium leguminosarum* bv. *trifolii* cultures were grown in TY or YM medium (Vincent 1970), supplemented where necessary with streptomycin (Str; 100 μg/ml) or neomycin (Neo; 200 μg/ml). *E. coli* cultures were grown in either TY or LB (Miller 1972) medium. Antibiotic concentrations used for *E. coli* were as follows: tetracycline (Tet; 15 μg/ml); kanamycin (Kan; 25 μg/ml); nalidixic acid (Nal; 25 μg/ml) and gentamycin (Gen; 25 μg/ml).

Crosses.

Crosses were carried out on TY medium by the patch plate method (Dixon et al. 1976). Transfer of pLAFR1 cosmids into rhizobia or *E. coli* was carried out by the triparental mating system (Ditta et al. 1980) using pRK2073 as the helper plasmid. *Rhizobium* transconjugants were selected on S20 medium containing streptomycin and tetracycline and then mass inoculated onto *L. pedunculatus* seedlings. Rhizobia were isolated from nodules, single colony purified, and the resident pLAFR1 cosmid then transferred by conjugation to *E. coli* HB101. To confirm that bacteria isolated from

```
RlnodX  ..........  .......MGP  SNEHSSGHRN  NFDLLRLFAA  CQVMFSHAWN   33
Sf6oct  ..........  ..........  .....MHKSN  CFDTARLVAA  MMVLVSHHYA   25
Saloct  ..........  ......MEKN  SFPISHEHSL  TMDYVKAFGM  IFVLVGHI..   32
Xamoct  VNAVTGASGT  SAPVQAAGAR  AFASGRSRDP  RIDATKAIAI  LLVVFCHA..   48
RlnolI  ......MLDH  IRAGAKGRGS  CPAGTNNRDL  SFDFAKAILI  TLVIIGHLLQ   44
                                                *          *  *

RlnodX  WLHLGDPLNG  TWVFDLLFSA  PGVA..IFFL  ISGFLVTDSY  IRSSSAASFF   81
Sf6oct  LSGQPEP...  .YLFGFE.SA  GGIAVIIFFS  ISGYLISKSA  IRSDSFIDFM   70
Saloct  ......N.ND  IFNV.YY..A  YLFHMPLFFF  IGGVLYKDTR  CITNFTAHVI   72
Xamoct  ......K.GV  PHGMTLF..A  YSFHVPLFFL  VSGWLAAGYA  SRTTSLLQTI   89
RlnolI  YLIYQGT.DA  FWLSPYFKSI  YMFHMPLFMA  ISGYLSSG.A  ILRKSFTQGV   92
                                            *     * *

RlnodX  VKRSLRIFPA  LFVNIAVMEL  ALLVTGGLNV  TG........  ..........   133
Sf6oct  AKRARRIFPA  L.VPCSILTY  FLF...GWIL  ND........  ..........    98
Saloct  KKQLPYLIVT  YLI...IGSI  ALLINVRYGI  HTGDAFSTGL  YETVKLAIKS   119
Xamoct  TKQARGLLLP  YVVFYLLGYV  YWLLTRNIGE  KAARWGSHPW  WEPI.VSMFT   138
RlnolI  GERAMQLLLP  MLFWCTLIW.  ........TL  KSAVIFPTKS  LTDTLLDLLT   133

RlnodX  ......ILQY  LFYFTVYILT  AARIWAVYFT  YEPYTMSGFY  GASDPSGV.L   156
Sf6oct  ........FS  AEYFSHDIVR  KTISSIFMSQ  APDADITSHL  IHAGING.SL   139
Saloct  NFHNNKMFLT  GWFLFAYIFV  SILSVIIIKS  IKRVVVSNAL  LLSVLVAISV   169
Xamoct  GVGPDLYVQP  PLWFLPVMLV  TVIGYVLLRR  WMPPLVIAA.  ..........   177
RlnolI  E......VIG  TYWF...IWA  AFISFILIR.  .....VLTTF  NRLSMWIISA   168
                    —*

RlnodX  WTLTVELTFY  LTLPMLLEIW  RRWKRAGALV  VAVAALGSWV  MAQHFNITDK   206
Sf6oct  WTLPLEFLCY  IITGVAVAL.  ..LKNGKAFI  VIQLVFVSLS  LIG..SVSEN   184
Saloct  LLITVSIT.Y  LSPQYILVKD  YKLNFICQVL  T.GMSFYIFG  .YVIRNQIYN   216
Xamoct  ..VAVVLA.W  FWMNWFPLQH  MRLFWGLDVL  PVSLCFYALG  ALLIHVSPYL   224
RlnolI  SAIAVAFA.P  ITLS......  .....ITPLL  KYTYPFYCLG  FLFAQPIGWQ   206

RlnodX  YNPFLSVTAG  PTFWIFSMGV  LARLYWHRVS  KIFEGKLLWW  LATHLAITWW   256
Sf6oct  RDVMFSI...  .PLWLYPLRG  LA.FFFGATM  AMYE.KSWNV  SNVKITVVSL   228
Saloct  LLNFYVFILL  TVIL.....Y  VS....KSYG  FSTQTIMSWS  YYPDGLIMS.   256
Xamoct  PTSLPGSALV  TVVLAALVAW  LA....GVNG  RIDVNMLEFG  RQHAVFLLSA   270
RlnolI  NGVIWRYKSI  FVVLFSIAAF  ICFLGWGKET  YAYNNLVLIH  DEQSAKRVFL   256

RlnodX  VAGTSAAFIS  INNAAPVDAF  RIAVLAGLVL  SAAHSLPRPN  LLRRQDLSYG   306
Sf6oct  LAMYAYASYG  KGIDYTMTCY  ILVSFSTIAI  CT..SVGDPL  VKGRFDYSYG   276
Saloct  VINALIGIYA  VFFISLLITR  GMKEIKLLKM  .....IGQNS  RA........   293
Xamoct  VA.GSLMVIC  A.......AR  MVQEWTWLQW  .....IGRNT  LL........   299
RlnolI  MFSGSLAASA  VRMQSMFQCW  RLVYSTRVAR  FVAVQLGQST  LL........   298

RlnodX  IYLYHMLVMH  TLIAIGWV..  .......GHW  WLWIVEPVGT  VALAALSWAL   347
Sf6oct  VYIYAFPVQQ  VVINTLHM..  .......GFY  PSMLLSAVTV  LFLSHLSWNL   317
Saloct  IMAYHLLVYV  ILDIIASILG  DYSLSGTDVY  DNHFITKWSV  PVYIALG.LL   342
Xamoct  ILCTHMLVFF  VLSGVAALAG  GFGGARPGLG  WAIFVTLFAL  VASVPLRWFL   349
RlnolI  LYLVQGAVFR  LMDLIQ..FG  EVWNLTTRIA  FATVLGVAIV  VIAMAIRSIA   346
                *

RlnodX  IEQPAMKLRT  SLVARRLSVA  ..........   367
Sf6oct  VEKRFLT.RS  S...PKLSLD  ..........   333
Saloct  .....LPLIF  SILKQKVIGK  IKFKRDFRIN   367
Xamoct  .....MRFAP  WTLGARPVSA  ..........   364
RlnolI  ...RNLGYVS  RIVVGAPPHP  SPLKSQSVIN   373
```

Fig. 5. Alignment of the deduced polypeptide sequence of NolL with other *O*-acetyl transferases, including NodX. The *O*-acetyl transferases include *Rhizobium leguminosarum* NodX (RlnodX; Davis et al. 1988, accession no. X07990), *Shigella flexneri* bacteriophage Sf6 (Sf6oct; Clark et al. 1991, accession no. X59553), *Salmonella enterica* (Saloct; Wang et al. 1992, accession no. X60666), *Xanthomonas campestris* (Xamoct; Y. S. Lin et al., unpublished results, accession no. X78451) and *Rhizobium loti* NolL (RlnolL accession no. U22899). Structurally similar amino acids are shown in bold and the asterisk indicates identity in all five sequences.

nodules were not revertants, resident cosmids were transferred back to strain PN233 and these transconjugants were inoculated onto *L. pedunculatus* seedlings.

Tn*5* mutagenesis of *nod* cosmids.

This was carried out by the method of de Bruijn and Lupski (1984) using phage λ467 as the source of Tn*5*. Transposon Tn*5* insertions in *nod* cosmids were selected by conjugating each cosmid into *E. coli* C2110 and selecting for $Kan^rNal^rTet^r$ transconjugants. Plasmid DNA was prepared from randomly selected colonies and digested with *Eco*RI to identify clones containing Tn*5* in the *R. loti* insert DNA rather than in vector. The positions of the Tn*5* insertions were then mapped by analyzing restriction enzyme digests on gels and where necessary by probing with appropriate gel-isolated fragments. Marker exchange of Tn*5* insertions into *R. loti* strains PN184 (NZP2037*str-1*) or PN4115 (NZP2213*str-1*) was carried out by plasmid incompatibility using pPH1JI (Ruvkun and Ausubel 1981). Total DNA was isolated from these Str^r Gen^r Tet^r colonies, digested with *Eco*RI and Southern blots of the digests probed with the appropriate *Eco*RI fragment to confirm that a double crossover had occurred.

Nodulation tests.

Nodulation tests were carried out using *Lotus pedunculatus* 'Grasslands Maku', *Lotus corniculatus* L., *Trifolium pratense* L., 'Grasslands Hamua', and *Leucaena leucocephala* (Lam.) de Wit as previously described (Scott and Ronson 1982). After 6 weeks (8 to 10 weeks for *L. leucocephala*) plants were examined for the presence (Nod^+) or absence (Nod^-) of nodules and for symptoms of nitrogen starvation by measuring the dry weight of the tops of both inoculated and uninoculated plants. No significant difference in dry weight between the test and control plants indicated an ineffective (Fix^-) symbiosis. In the case of *L. leucocephala* acetylene reduction assays were carried out to determine whether nodules were

Table 3. Bacterial strains and plasmids

Strains, plasmids, phages	Relevant characteristics	Source or reference
Strains		
Rhizobium loti		
NZP2037	Nod^+Fix^+ (*Lotus pedunculatus, Lotus corniculatus,* and *Leucaena leucocephala*)	DSIR Culture Collection
PN184	NZP2037 *str-1*	Chua et al. 1985
PN233	PN184 *nodC*::Tn*5*	Chua et al. 1985
NZP2213	Nod^+Fix^+ (*Lotus corniculatus*), Nod^+Fix^- (*Lotus pedunculatus* and *Leucaena leucocephala*)	DSIR Culture Collection
PN4115	NZP2213 *str-1*	Jones et al. 1987
R. leguminosarum bv. *trifolii*		
ANU843	Nod^+Fix^+	Scott et al. 1982
ANU252	*nodA*::Tn*5*	Djordjevic et al. 1985
ANU249	*nodB*::Tn*5*	Djordjevic et al. 1985
ANU277	*nodC*::Tn*5*	Djordjevic et al. 1985
ANU851	*nodD*::Tn*5*	Scott et al. 1982
PN100	Nod^+Fix^+ *str-1 rif-1*	Scott and Ronson 1982
Escherichia coli		
JM101	*supE, thi,* Δ(*lac-proAB*), [F', *traD36, proAB, lacI*^{q} *lac* ZΔM15]	Yanisch-Perron et al. 1985
HB101	*pro leu thi gal lacY recA str hsdD hsdM*	Boyer and Roulland-Dussoix 1969
C2110	*polA* Nal^r	Leong et al. 1982
PN232	HB101/pPN306	Scott et al. 1985
PN457	HB101/pPN25	Scott et al. 1985
PN464	HB101/pPN366	This study
PN467	HB101/pPN369	Scott et al. 1985
PN600	HB101/pPN26	Scott et al. 1985
PN612	HB101/pPN377	This study
PN623	HB101/pPN381	This study
PN625	HB101/pPN383	This study
PN627	HB101/pPN385	This study
PN629	HB101/pPN387	This study
PN630	HB101/pPN388	This study
PN631	HB101/pPN389	This study
Plasmids		
pRK2073	Kan^s derivative of pRK2013	Leong et al. 1982
pPH1JI	Gen^r IncP	Beringer et al. 1978
pBR328	$Amp^rTet^rCam^r$	Bolivar et al. 1977
pLAFR1	λcos derivative of pRK290	Friedman et al. 1982
pRt572	7.2-kb *Eco*RI *nod* fragment from ANU843 cloned in pBR328	Schofield et al. 1983
pUC18	Amp^r	Norrander et al. 1983
pUC19	Amp^r	Norrander et al. 1983
pUC118	Amp^r	Vieira and Messing 1987
pUC119	Amp^r	Vieira and Messing 1987
Phages		
λ467		de Bruijn and Lupski 1984
M13mp18		Norrander et al. 1983
M13mp19		Norrander et al. 1983

Fix$^+$ or Fix$^-$, instead of measuring the dry weight of tops, because of the very slow growth rate of these plants. In all nodulation tests the wild-type strains were included for comparison. Bacteria were isolated from nodules as previously described (Scott and Ronson 1982).

Molecular biology methods.

All DNA manipulation procedures, including DNA sequencing methods, have been previously described (Scott et al. 1985; Ward et al. 1989). Primers used for DNA sequencing, other than the universal M13 17mer (New England Biolabs, Beverly, MA), included a Tn*5* sequencing primer (5′-CGTTCAGGACGCTACTT-3′) previously described by Schofield and Watson (1986) and primers

BS4 (5′-AAGAACTCAGAGAGTTCGAG-3′),
BS5 (5′-CGTCGCTCGACTCTGAGACG-3′),
BS6 (5′-GCACCCGATTTTGCGCCAGC-3′),
BS7 (5′-GCCTAACAGTCGGGCCATCC-3′),
BS8 (5′-ATGAGTTGCAGCTTTCCGAT-3′),
BS11 (5′-GATTTTGTCCGCTCGGTTGC-3′),

used for generating complete double-stranded sequence as described in Results.

Computer analysis.

DNA sequences were entered, assembled, and analyzed on a VAX using the Genetics Computer Group, Inc., (GCG) package (Devereux et al. 1984).

ACKNOWLEDGMENTS

We are grateful to Barry Rolfe and John Watson for providing strains and plasmids, respectively. This research was supported from block grant funding to DSIR and Massey University. Julie Collins-Emerson was supported by a Ph.D. fellowship from Massey University.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Appelbaum, E. R., Thompson, D. V., Idler, K., and Chartrain, N. 1988. *Rhizobium japonicum* USDA 191 has two *nodD* genes that differ in primary structure and function. J. Bacteriol. 170:12-20.

Atkinson, E. M., Palcic, M. M., Hindsgaul, O., and Long, S. R. 1994. Biosynthesis of *Rhizobium meliloti* lipooligosaccharide Nod factors: NodA is required for an *N*-acyltransferase activity. Proc. Natl. Acad. Sci. USA. 91:8418-8422.

Beringer, J. E. 1974. R factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. 84:118-198.

Beringer, J. E., Beynon, J. L., Buchanan-Wollaston, A. V., and Johnston, A. W. B. 1978. Transfer of the drug-resistant transposon Tn*5* to *Rhizobium*. Nature 276:633-634.

Bloemberg, G. V., Thomas-Oates, J. E., Lugtenberg, B. J. J., and Spaink, H. P. 1994. Nodulation protein NodL of *Rhizobium leguminosarum O*-acetylates lipo-oligosaccharides, chitin fragments and *N*-acetylglucosamine *in vitro*. Mol. Microbiol. 11:793-804.

Bolivar, F., Rodriguez, R., Greene, P. J., Betlach, M., Heyneker, H. L., Boyer, H. W., Crosa, J., and Falkow, S. 1977. Construction and characterisation of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95-113.

Boyer, H. W., and Roulland-Dussoix, D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. 41:459-472.

Bulawa, C. E., and Wasco, W. 1991 Chitin and nodulation. Nature 353:710

Chua K.-Y., Pankhurst, C. E., Macdonald, P. E., Hopcroft, D. H., Jarvis, B. D. W., and Scott, D. B. 1985. Isolation and characterisation of Tn*5*-induced symbiotic mutants of *Rhizobium loti*. J. Bacteriol. 162:335-343.

Clark, C. A., Beltrame, J., and Manning P. A. 1991. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. Gene 107:43-52.

Collins-Emerson, J. M., Terzaghi, E. A., and Scott, D. B. 1990. Nucleotide sequence of *Rhizobium loti nod*C. Nucleic Acids Res. 18:6690.

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

Davis, E. O., and Johnston, A. W. B. 1990. Analysis of three *nod*D genes in *Rhizobium leguminosarum* biovar *phaseoli*; *nodD*1 is preceded by *nolE*, a gene whose product is secreted from the cytoplasm. Mol. Microbiol. 4:921-932.

de Bruijn, F. J., and Lupski, J. R. 1984. The use of transposon Tn*5* mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - a review. Gene 27:131-149.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-395.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 1980. Broad host range DNA cloning system for Gram negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA 77: 7347-7351.

Dixon, R., Cannon, F., and Kondorosi, A. 1976. Construction of a P plasmid carrying nitrogen fixation genes from *Klebsiella pneumoniae*. Nature 260:268-271.

Djordjevic, M. A., Schofield, P. R., and Rolfe, B. G. 1985. Tn*5* mutagenesis of *Rhizobium trifolii* host-specific nodulation genes result in mutants with altered host-range ability. Mol. Gen. Genet. 200: 463-471.

Downie, J. A. 1994. Signalling strategies for nodulation of legumes by rhizobia. Trends Microbiol. 2:318-324.

Egelhoff, T. T., Fisher, R. F., Jacobs, T. W., Mulligan, J. T., and Long, S. R. 1985. Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. DNA 4:241-248.

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

Firmin, J. L., Wilson, K. E., Carlson, R. W., Davies, A. E., and Downie, J. A. 1993. Resistance to nodulation of cv. Afghanistan peas is overcome by *nodX*, which mediates an *O*-acetylation of the *Rhizobium leguminosarum* lipo-oligosaccharide nodulation factor. Mol. Microbiol. 10:351-360.

Firmin, J. L., Wilson, K. E., Rossen, L., and Johnston, A. W. B. 1986. Flavonoid activation of nodulation genes in *Rhizobium* reversed by other compounds present in plants. Nature 324:90-92.

Fisher, R. F., Egelhoff, T. T., Mulligan, J. T., and Long, S. R. 1988. Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. Genes & Dev. 2:282-293.

Friedman, A. M., Long, S. R., Brown, S. E., Buikema, W. J., and Ausubel, F. M. 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene 18:298-296.

Goethals, K., Van Montagu, M., and Holsters, M. 1992. Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. Proc. Natl. Acad. Sci. USA 89:1646-1650.

Göttfert, M., Horvath, B., Kondorosi, E., Putnoky, P., Rodriguez-Quinones, R., and Kondorosi, A. 1986. At least two *nodD* genes are necessary for efficient nodulation on alfalfa by *Rhizobium meliloti*. J. Mol. Biol. 191:411-420

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

Györgypal, Z., Iyer, N., and Kondorosi, A. 1988. Three regulatory *nodD* alleles of diverged flavonoid-specificity are involved in host-dependent nodulation by *Rhizobium meliloti*. Mol. Gen. Genet. 212:85-92.

Györgypal, Z., Kondorosi, E., and Kondorosi, A. 1991. Diverse signal

sensitivity of NodD protein homologs from narrow and broad host range rhizobia. Mol. Plant-Microbe Interact. 4:356-364.

Handberg, K., and Stougaard, J. 1992. *Lotus japonicus* an autogamous, diploid legume species for classical and molecular genetics. Plant J. 2:487-496.

Henikoff, S., Haughn, G. W., Calvo, J. M., and Wallace, J. C. 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA 85:6602-6606.

Hong, G.-F., Burn, J. E., and Johnston, A. W. B. 1987. Evidence that DNA involved in the expression of nodulation (*nod*) genes in *Rhizobium* binds to the product of the regulatory gene *nodD*. Nucleic Acids Res. 15:9677-9691.

Honma, M. A., and Ausubel, F. M. 1987. *Rhizobium meliloti* has three functional copies of the *nodD* symbiotic regulatory gene. Proc. Natl. Acad. Sci. USA 84:8558-8562.

Horvath, B., Bachem, C. W. B., Schell, J., and Kondorosi, A. 1987. Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the *nodD* gene product. EMBO J. 6: 841-848.

Innes, R. W., Kuempel, P. L., Plazinski, J., Canter-Cremers, H., Rolfe, B. G., and Djordjevic, M. A. 1985. Plant factors induce expression of nodulation and host-range genes in *Rhizobium trifolii*. Mol. Gen. Genet. 201:426-432.

Jacobs, T. W., Egelhoff, T. T., and Long, S. R. 1985. Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nodC*. J. Bacteriol. 162:469-476.

John, M., Röhrig, H., Schmidt, J., Wieneke, U., and Schell, J. 1993. *Rhizobium* NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. Proc. Natl. Acad. Sci. USA 90:625-629.

Jones, W. T., MacDonald, P. E., Jones, S. D., and Pankhurst, C. E. 1987. Peptidoglycan-bound polysaccharide associated with resistance of *Rhizobium loti* strain NZP2037 to *Lotus pedunculatus* root flavolan. J. Gen. Microbiol. 133:2617-2629.

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

Leong, S. A., Ditta, G. S., and Helinski, D. R. 1982. Heme biosynthesis in *Rhizobium*. Identification of a cloned gene coding for δ-aminolevulinic acid synthesis from *Rhizobium meliloti*. J. Biol. Chem. 257: 8724-8730.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J. C., and Dénarié, J. 1990. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glycosamine oligosaccharide signal. Nature 344:781-784.

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

Long, S. R., Buikema, W. J., and Ausubel, F. M. 1982. Cloning *Rhizobium meliloti* nodulation genes by direct complementation of Nod$^-$ mutants. Nature 298:485-488.

López-Lara, I. M., van den Berg J. D. J., Thomas-Oates, J. E., Glushka, J., Lugtenberg, B. J. J., and Spaink, H. P. 1995. Structural identification of the lipo-chitin oligosaccharide nodulation signals of *Rhizobium loti*. Mol. Microbiol. 15:627-638.

McKay, I. A., and Djordjevic, M. A. 1993. Production and excretion of Nod metabolites by *Rhizobium leguminosarum* bv. *trifolii* are disrupted by the same environmental factors that reduce nodulation in the field. Appl. Environ. Microbiol. 59:3385-3392.

Mergaert, P., van Montagu, M., Promé, J.-C. and Holsters, M. 1993. Three unusual modifications, a D-arabinosyl, an N-methyl, and a carbamoyl group, are present on the Nod factors of *Azorhizobium caulinodans* strain ORS571. Proc. Natl. Acad. Sci. USA 90:1551-1555.

Miller, J. M., ed. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mulligan, J. T., and Long, S. R. 1985. Induction of *Rhizobium meliloti nodC* expression by plant exudate requires *nodD*. Proc. Natl. Acad. Sci USA 82:6609-6613.

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

Pankhurst, C. E., Craig, A. S., and Jones, W. T. 1979. Effectiveness of *Lotus* root nodules. I. Morphology and flavolan content of nodules formed on *Lotus pedunculatus* by fast-growing *Lotus* rhizobia. J. Exp. Bot. 30:1085-1093.

Pankhurst, C. E., Hopcroft, D. H., and Jones, W. T. 1987. Comparative morphology and flavolan content of *Rhizobium loti* induced effective and ineffective root nodules on *Lotus* species, *Leuceana leucocephala*, *Carmichaelia flagelliformis*, *Ornithopus sativus*, and *Clianthus puniceus*. Can. J. Bot. 65:2676-2685.

Peters, K. N., Frost, J. W., and Long, S. R. 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science 233:977-980.

Poupot, R., Martínez-Romero, E., and Promé, J.-C. 1993. Nodulation factors from *Rhizobium tropici* are sulfated or nonsulfated chitopentasaccharides containing an *N*-methyl-*N*-acylglucosaminyl terminus. Biochemistry 32:10430-10435.

Price, N. P. J., Relic, B., Talmont., F., Lewin, A., Promé, D., Pueppke, S. G., Maillet, F., Dénarié, J., Promé, J.-C. and Broughton, W. J. 1992. Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are *O*-acetylated or sulphated. Mol. Microbiol. 6:3575-3584.

Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., and Rolfe, B. G. 1986. Flavones induce expression of nodulation genes in *Rhizobium*. Nature 323:632-635.

Roche, P., Debellé, F, Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Dénarié, J., and Promé, J.-C. 1991. Molecular basis of symbiotic host specificity in *Rhizobium meliloti* : *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. Cell 67:1131-1143.

Röhrig, H., Schmidt, J., Wieneke, U., Kondorosi, E., Barlier, I., Schell, J., and John, M. 1994. Biosynthesis of lipooligosaccharide nodulation factors: *Rhizobium* NodA protein is involved in N-acylation of the chitooligosaccharide backbone Proc. Natl. Acad. Sci. USA 91:3122-3126.

Rossen, L., Shearman, C. A., Johnston, A. W. B., and Downie, J. A. 1985. The *nodD* gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nodA, B, C* genes. EMBO J. 4:3369-3373.

Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A., and Kondorosi, A. 1986. Conservation of extended promoter regions of nodulation genes in *Rhizobium*. Proc. Natl. Acad. Sci. USA 83:1757-1761.

Ruvkun, G. B., and Ausubel, F. M. 1981. A general method for site-directed mutagenesis in prokaryotes. Nature 289:85-88.

Schofield, P. R., Djordevic, M. A., Rolfe, B. G., Shine, J., and Watson, J. M. 1983. A molecular linkage map of nitrogenase and nodulation genes in *Rhizobium trifolii*. Mol. Gen. Genet. 192:459-465.

Schofield, P. R., and Watson, J. M. 1986. DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. Nucleic Acids Res. 14: 2891-2903.

Schultze, M., Quiclet-Sire, B., Kondorosi, E., Virelizier, H., Glushka, J. N., Endre, G., Géro, S. D., and Kondorosi, A. 1992. *Rhizobium meliloti* produces a family of sulfated lipooligosaccharides exhibiting different degrees of plant host specificity Proc. Natl. Acad. Sci. USA 89: 192-196.

Schwedock, J., and Long, S. R. 1990. ATP sulphurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. Nature 348:644-647.

Scott, D. B., Chua, K.-Y., Jarvis, B. D. W., and Pankhurst, C. E. 1985. Molecular cloning of a nodulation gene from fast- and slow-growing strains of *Lotus* rhizobia. Mol. Gen. Genet. 201:43-50.

Scott, D. B., and Ronson, C. W. 1982. Identification and mobilisation by cointegrate formation of a nodulation plasmid in *Rhizobium trifolii*. J. Bacteriol. 151:36-43.

Scott, K. F., Hughes, J. E., Gresshoff, P. M., Beringer, J. E., Rolfe, B. G., and Shine, J. 1982. Molecular cloning of *Rhizobium trifolii* genes involved in symbiotic nitrogen fixation. J. Mol. Appl. Genet. 1:315-326.

Segovia, L., Young, J. P. W., and Martínez-Romero, E. 1993. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* Type I strains as *Rhizobium etli sp. nov.* Int. J. Syst. Bacteriol. 43:374-377.

Spaink, H. P., Sheeley, D. M., van Brussel, A. A. N., Glushka, J., York, W. S., Tak, T., Geiger, O., Kennedy, E. P., Reinhold, V. N., and Lugtenberg, B. J. J. 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. Nature 354:125-130.

Spaink, H. P., Wijffelman, C. A., Pees, E., Okker, R. J. H., and Lugtenberg, B. J. J. 1987. *Rhizobium* nodulation gene *nodD* as a determinant of host specificity. Nature 328:337-340.

Stacey, G., Luka, S., Sanjuan, J., Banfalvi, Z., Nieuwkoop, A. J., Chun, J. Y., Forsberg, L. S., and Carlson, R. 1994. *NodZ*, a unique host-specific nodulation gene, is involved in the fucosylation of the lipooligosaccharide nodulation signal of *Bradyrhizobium japonicum*. J. Bacteriol. 176:620-633.

Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., de Billy, F., Promé, J.-C., and Dénarié, J. 1991. Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. Nature 351:670-673.

Vázquez, M., Dávalos, A., de las Peñas, A., Sánchez, F., and Quinto, C. 1991. Novel organization of the common nodulation genes in *Rhizobium leguminosarum* bv *phaseoli* strains. J. Bacteriol. 173:1250-1258.

Vieira, J., and Messing, J. 1987. Production of single-stranded plasmid DNA. Methods Enzymol. 153:3-11.

Vincent, J. M. 1970. A Manual for the Practical Study of the Root-Nodule Bacteria. Blackwell Scientific Publications, Oxford.

Wang, L., Romana, L. K., and Reeves, P R. 1992. Molecular analysis of a *Salmonella enterica* Group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics 130:429-443.

Ward, L. J. H., Rockman, E. S., Ball, P., Jarvis, B. D. W., and Scott, D. B. 1989. Isolation and characterization of a *Rhizobium loti* gene required for effective nodulation of *Lotus pedunculatus*. Mol. Plant-Microbe Interact. 2:224-232.

Yanisch-Perron, C., Vieira, J., and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

Young, C., Collins-Emerson, J. M., Terzaghi, E. A., and Scott, D. B. 1990. Nucleotide sequence of *Rhizobium loti nodI*. Nucleic Acids Res. 10:6691.

Young, J. P. W., and Johnston A. W. B. 1989. The evolution of specificity in the legume-Rhizobium symbiosis. Trends Ecol. & Evol. 4:341-349.