MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 4, No. 5, pp. 521-529, 1991

Repetitive Sequences with Homology to *Bradyrhizobium japonicum* DNA and the T-DNA of *Agrobacterium rhizogenes* Are Closely Linked to *nodABC* of *Rhizobium fredii* USDA257

Hari B. Krishnan and Steven G. Pueppke

Department of Plant Pathology, University of Missouri, Columbia 65211 U.S.A.
Received 21 March 1991. Accepted 17 May 1991.

We have detected strong homology between a 9.2-kb *Eco*RI restriction fragment from *Rhizobium fredii* USDA257 that contains *nodABC* and eight additional *Eco*RI fragments in DNA digests from this organism. A series of repetitive sequences responsible for this hybridization lies within a 0.95-kb *Hin*dIII/*Sal*I subfragment about 1-kb upstream of *nodA*. This subfragment also hybridizes to multiple restriction fragments from nine other strains of *R. fredii*, but only one is common to all strains. The 0.95-kb subfragment does not hybridize to genomic DNA from 17 other strains of fast-growing rhizobia, but there is weak homology to two fragments from *Rhizobium* sp. strain NGR234. We sequenced 2,432 base pairs (bp) of the region encompassing the repetitive sequences. It contains 65 separate 8- to 11-bp inverted and direct repeats, as well as two large open reading frames (ORFs) that overlap on opposite strands. ORF1 reads in the same direction as *nodABC*, contains 1,071 bp, and encodes a 40.6-kD protein. It has 74% sequence homology to an ORF within the T-DNA of *Agrobacterium rhizogenes* and similar homology to a series of repetitive sequences from *Bradyrhizobium japonicum*. ORF2 (981 bp) reads in the opposite direction, encodes a 34.7-kD protein, and has partial identity with a second ORF from *A. rhizogenes*. We could detect no poly(A)$^+$ nodule transcripts with homology to ORF1 and ORF2. The eight sets of repetitive sequences found in other *Eco*RI fragments of the genome were cloned from USDA257 on separate cosmids. Some of these cosmids appear to overlap, and two have fragments with homology to *nifKDH*.

Additional keyword: common *nod* genes.

Rhizobium and the allied genera *Azorhizobium* and *Bradyrhizobium* are gram-negative soil-dwelling bacteria that fix nitrogen in symbiosis with legumes. The genetic basis for plant infection, host range, and nitrogen fixation is under intensive study and has led to the description of genes that have both general and highly specific functions in the process (Martinez *et al.* 1990). Random transposon mutagenesis followed by screening for altered symbiotic phenotypes on plants has allowed simultaneous identification and physical tagging of both nodulation (*nod*) and nitrogen fixation (*fix*) genes (Simon and Priefer 1989).

One recurring and initially puzzling observation in these studies was the appearance of symbiotic mutants having the transposon inserted into some neutral site unrelated to the defect in nodulation (Ruvkun *et al.* 1982; Kondorosi *et al.* 1984; Dusha *et al.* 1987; Wheatcroft and Watson 1987). Such mutants are especially common in *Rhizobium meliloti* Dangeard and have led to the accidental discovery of the endogenous insertion sequences IS*Rm1* (Ruvkun *et al.* 1982) and IS*Rm2* (Dusha *et al.* 1987). These elements, which normally are present in 10 or fewer copies per cell (Ruvkun *et al.* 1982; Dusha *et al.* 1987; Wheatcroft and Watson 1988; Labes and Simon 1990), may be activated during transposon mutagenesis and preferentially inserted into *nod*/*nif* regions. IS*Rm3* was discovered during a search of *R. meliloti* plasmids (Wheatcroft and Laberge 1991), and another such element was discovered in *R. lupini*, where it transposes into plasmid RP4 at high frequency (Priefer *et al.* 1981). In *Bradyrhizobium japonicum* (Buchanan) Jordan, several families of insertion sequencelike elements are known to be clustered in *nod*/*nif* regions of the genome, where they apparently promote deletions and other genetic rearrangements (Kaluza *et al.* 1985; Hahn and Hennecke 1987a, 1987b). The presence of reiterated DNA sequences also has been implicated in genetic rearrangements that occur in other strains of rhizobia (Flores *et al.* 1987).

In addition to known or putative transposable elements, the genomes of several rhizobia contain other types of repeated sequences. These include reiterations of the *nifHKD* promoter region (Better *et al.* 1983; Watson and Schofield 1985), as well as a series of poorly characterized reiterations in *R. fredii* Scholla and Elkan (Prakash and Atherly 1984; Masterson and Atherly 1986). Intact *nod* and *nif* genes also may be present in several copies per genome. Examples include *nodD* (Young and Johnston 1989), which regulates root hair curling and infection, and *nifH* (Quinto *et al.* 1982, 1985), which encodes nitrogenase reductase.

We are particularly interested in strain USDA257 of *R. fredii*. This organism has a broad host range for legume species, yet exhibits cultivar specificity in nodulation of soybean (Keyser *et al.* 1982; Heron and Pueppke 1984; Scholla *et al.* 1984). We recently cloned and sequenced a number of *nod* genes from USDA257, including *nodABC*, an operon that functions in initial root hair curling (Krishnan and Pueppke 1991). *nodABC* is wholly contained

Address correspondence to Steven G. Pueppke.

Nucleotide and/or amino acid sequence data is to be submitted to GenBank, EMBL, and DDBJ as accession number J03701.

© 1991 The American Phytopathological Society

within a 9.2-kb *Eco*RI fragment from the Sym plasmid of USDA257 (Heron *et al.* 1989; Krishnan and Pueppke 1991). When we used this fragment to probe *Eco*RI-digested genomic DNA of USDA257, we were surprised to detect eight additional restriction fragments with strong homology. We report here the sequencing and organization of the *nodABC*-linked repetitive sequences. This region contains overlapping ORFs with significant homology to the T-DNA of *Agrobacterium rhizogenes* (Riker *et al.*) Conn, as well as to a family of repeated sequencelike insertion sequences from *B. japonicum*. We also provide data on other copies of the reiterations in *R. fredii* as well as a search for these sequences in other rhizobia.

MATERIALS AND METHODS

Strains and plasmids. Wild-type strains of *R. fredii*, including USDA257, originally were from the collection of the U.S. Department of Agriculture and have been described (Heron and Pueppke 1984). Nod$^-$ transposon-mutants 257B3, 257F3, and 257M5 were prepared and analyzed in our laboratory (Heron *et al.* 1989). Strains of *R. meliloti* and *R. leguminosarum* Frank bv. *phaseoli* Jordan and *trifolii* Jordan prefixed USDA were from the U.S. Department of Agriculture. D. H. Hubbell (University of Florida) provided *R. leguminosarum* bv. *trifolii* strain BAL and *viciae* Jordan strain 3HOq1. T. A. Lie (Agricultural University of Wageningen) supplied *R. leguminosarum* bv. *viciae* strains PRE, PF_2, ANKK319, and TOM. C. P. Vance (University of Minnesota) donated *R. meliloti* strains 102F29, 102F51, and 102F65; and W. J. Broughton (University of Geneva) provided *R. meliloti* RCR2011 and *Rhizobium* sp. NGR234. DNA libraries and clones from USDA257 were prepared in our laboratory (Heron *et al.* 1989; Krishnan and Pueppke 1991); plasmids are listed in Table 1. All bacteria were maintained in 7.5% glycerol at −70° C. *Escherichia coli* was cultured in yeast extract-mannitol (YEM) medium (Vincent 1970) at 37° C. Antibiotics were supplied as necessary at the following concentrations: carbenicillin (100 μg/ml), kanamycin (100 μg/ml), and tetracycline (10 μg/ml).

General molecular methods. Cloning, plasmid isolation, restriction, and Southern hybridizations followed standard procedures (Maniatis *et al.* 1982). Genomic DNA was isolated and purified as described (Jagadish and Szalay 1984), and DNA probes incorporating [^{32}P]-dCTP were synthesized by random priming with Klenow enzyme (USB Biochemicals, Cleveland, OH). Southern hybridizations proceeded overnight at 68° C in 10× Denhardt's solution, 6× SSC (1× SSC = 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0) containing 0.1 mg of salmon sperm DNA per milliliter. Unless stated otherwise, filters were washed in a solution of 0.1% sodium dodecyl sulfate in 0.3× SSC at 68° C prior to autoradiography. Procedures for colony hybridizations have been described (Heron *et al.* 1989).

Sequence analysis. A total of 2,432 bp were sequenced by the dideoxy-chain termination procedure (Sanger *et al.* 1977). This region corresponds to the entire 2,311-bp insert of pHBK312 and an additional 121-bp extending to the right of the *Hin*dIII site (Fig. 1). The strategy involved generation of a series of subclones of pHBK312. Sequencing was initiated at the restriction sites, and the sequences were extended with the aid of a series of oligonucleotide primers. The entire sequence was verified by complete analysis of both strands. Computerized manipulation of the sequence was by means of the EuGene software package, obtained courtesy of the DNA Core Facility, University of Missouri.

Northern hybridizations. Nodules were induced by *R. fredii* 257DH4 on supernodulating Williams soybean. This combination is particularly efficient in nodule production (Balatti and Pueppke 1990). Total RNA was isolated from nodules harvested 20 days after inoculation as described (Rochester *et al.* 1986), including two precipitations with 2 M LiCl. Poly(A)$^+$ RNA was enriched by the method of Aviv and Leder (1972). RNA fractions were resolved on 1.5% agarose gels containing formaldehyde and transferred by capillary blotting to nitrocellulose. The gels were hybridized by standard procedures (Maniatis *et al.* 1982), with the 0.95-kb insert of pHBK311 as probe. Leghemoglobin clone pLB23 (Fuller and Verma 1984) was used as a positive control probe in these experiments.

RESULTS

Localization of repetitive sequences in USDA257. *nodABC* of *R. fredii* USDA257 is included in a 9.2-kb

Table 1. Plasmids used in this study

Plasmid	Vector	Insert
pHBK301	pGEM7Zf (+); Ap^R (Promega Biotech)[a]	9.2-kb *Eco*RI fragment containing *nodABC* and repetitive sequences
pHBK310	pGEM7Zf (+); Ap^R (Promega Biotech)	1.7-kb *Bam*HI fragment containing most of *nodABC*
pHBK311	pTZ19u; Ap^R (USB Biochemicals)[b]	0.95-kb *Hin*dIII/*Sal*I fragment containing repetitive sequences
pHBK312	pGEM7Zf (+); Ap^R (Promega Biotech)	2.3-kb *Hin*dIII/*Eco*RI fragment containing all of ORF1 and ORF2

[a] Madison, WI.
[b] Cleveland, OH.

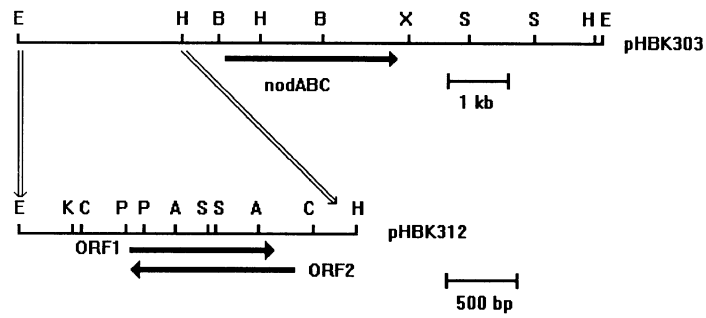

Fig. 1. Coordinated physical and genetic map of the 9.2-kb *nodABC*-containing *Eco*RI fragment of *Rhizobium fredii* USDA257. Restriction sites used for mapping the fragment are indicated. The thick arrow indicates the position of *nodABC* and its direction of transcription. The lower expanded map corresponds to the sequenced region containing repetitive sequences. The thick arrows below this portion of the map indicate the positions of ORF1 and ORF2 and the directions of their transcription. Abbreviations for restriction enzymes: A, *Ava*I; B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sal*I; X, *Xho*I.

*Eco*RI fragment from total genomic DNA (Krishnan and Pueppke 1991) (Fig. 1). When the 9.2-kb fragment was used to probe *Eco*RI-digested genomic DNA from USDA257, eight additional intensely hybridizing bands were detected (Fig. 2). They range in size from 16 to 1.3 kb and are most easily resolved in mutant 257B3, which has a deletion in the *nodABC* region (Heron *et al.* 1989; Krishnan and Pueppke 1991). All nine bands were present in the other Nod$^-$ mutants (Fig. 2).

To determine the extent of the homology common to these sequences, we isolated a series of subfragments from pHBK301 and used them to probe genomic DNA. Three subfragments (the rightmost 3.9-kb *Eco*RI/*Bam*HI fragment; the adjacent 1.7-kb *Bam*HI fragment; a 1.3-kb *Bam*HI/*Hin*dIII fragment that is internal to *nodABC*) hybridized only to the 9.2-kb fragment in *Eco*RI digests of genomic DNA (Fig. 3, lanes A–C). Together, these fragments spanned nearly three-quarters of the 9.2-kb fragment and included all of *nodABC*. In contrast, the 2.3-kb *Eco*RI/*Hin*dIII subfragment, which mapped to the left side of the 9.2-kb fragment (Fig. 1), hybridized to all nine genomic fragments. By using progressively smaller subfragments as probes, we ultimately localized the *nodABC*-linked repetitive sequences to a 0.95-kb *Hin*dIII/*Sal*I fragment that lies about 1 kb upstream from *nodA*. The adjacent 0.96-kb *Sal*I/*Kpn*I subfragment (note that a small intervening *Sal*I fragment was lost during digestion) hybridized with all but one fragment.

We retrieved eight of the nine repetitive sequence-containing fragments on separate cosmids from a DNA library of DNA from USDA257 (Fig. 4). The ninth and smallest fragment was linked to the 1.6-kb fragment on cosmid pRFRS9. Preliminary analysis of these and other cosmids confirmed additional linkages among the repetitive

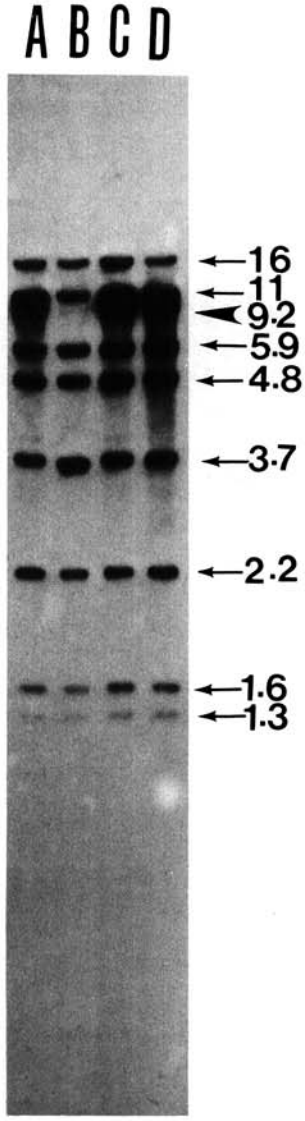

Fig. 2. Repetitive sequences in the genome of *Rhizobium fredii* USDA257 and mutants derived from it. A 9.2-kb *Eco*RI fragment, which contains *nodABC*, was hybridized to *Eco*RI-digested genomic DNA of (A) USDA257; (B) 257B3; (C) 257F3; (D) 257M5. The 9.2-kb fragment is identified with an arrowhead; other fragments are identified with arrows (sizes in kilobases).

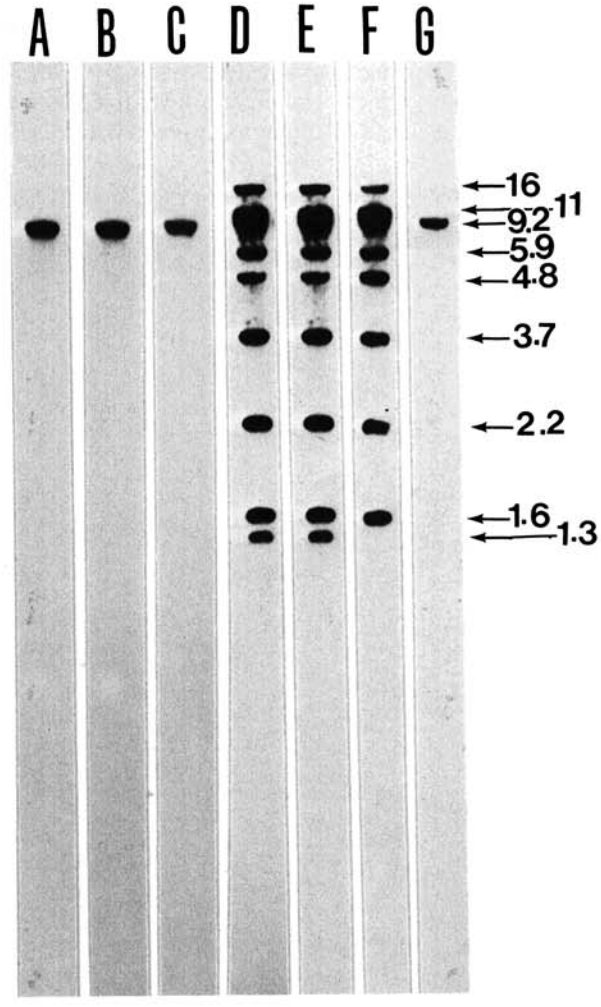

Fig. 3. Localization of repetitive sequences on the 9.2-kb *Eco*RI fragment. Genomic DNA of USDA257 was digested with *Eco*RI and blots were probed with internal subfragments of the 9.2-kb fragment (see Figure 1) as follows: (A) the 3.9-kb *Eco*RI/*Bam*HI fragment; (B) the 1.7-kb *Bam*HI fragment; (C) the 1.3-kb *Bam*HI/*Hin*dIII fragment; (D) the 2.3-kb *Hin*dIII/*Eco*RI fragment; (E) the 0.95-kb *Hin*dIII/*Sal*I fragment; (F) the 0.96-kb *Sal*I/*Kpn*I fragment; (G) the 0.37-kb *Kpn*I/*Eco*RI fragment. Arrows mark the sizes of the hybridizing fragments (sizes in kilobases).

sequences. For example, we retrieved an additional cosmid that contained both the original 9.2-kb *Eco*RI fragment and the 11-kb fragment. Using pSA30 (Cannon *et al.* 1979) as a probe for *nifKDH*, we also localized *nif* genes on a pair of *Eco*RI fragments, both of which are on cosmids pRFRS2 and pRFRS4 (data not shown). These observations confirmed the existence of *nif-nod* linkage in USDA257 and its physical association with repetitive sequences.

Distribution of the repetitive sequences in *R. fredii* and other rhizobia. Figure 5A shows that the repetitive sequences are widely distributed in *R. fredii*. All nine tested strains contained multiple *Eco*RI fragments with homology to the 0.95-kb *Hin*dIII/*Sal*I fragment, and although many of these fragments were large and overlapping, each strain appeared to contain from six to 11 hybridizing bands. The resulting polymorphisms were strain-specific, except those of USDA205 and USDA206, which were identical. One hybridizing fragment of 1.6 kb was present in all strains. Several fragments, however, including those of 5.9 and 2.2 kb, were diagnostic for USDA257 (Fig. 5A). The band heterogeneity of the repetitive sequences contrasted strikingly with the *nodABC* homology, which was invariant and represented by a single 9.2-kb *Eco*RI fragment in each strain (Fig. 5B).

Figure 6 extends the search for repetitive sequences to a series of 19 additional strains representing *R. meliloti*, *R. l.* bv. *phaseoli*, *trifolii*, and *viciae*, as well as broad host range *Rhizobium* sp. NGR234. With the exception of the latter, none of these strains had any fragments with detectable homology to the 0.95-kb *Hin*dIII/*Sal*I fragment from USDA257. Strain NGR234, however, contained two weakly hybridizing fragments of about 3 and 7 kb. As a positive control, DNA digests from each strain also were probed with the 1.7-kb *Bam*HI fragment of USDA257, which was internal to *nodABC*. Each strain possessed a single hybridizing fragment, and the fragments were variable in size. *R. l.* bv. *trifolii* USDA2717 (lane I), the single exception to this pattern, contained two large hybridizing fragments, suggesting that *nodABC* itself may be reiterated or fragmented in this strain.

Sequence analysis of the *nodABC*-linked repetitive sequences. To precisely identify the *nodABC*-linked repetitive sequences, we sequenced a 2,432-bp segment of

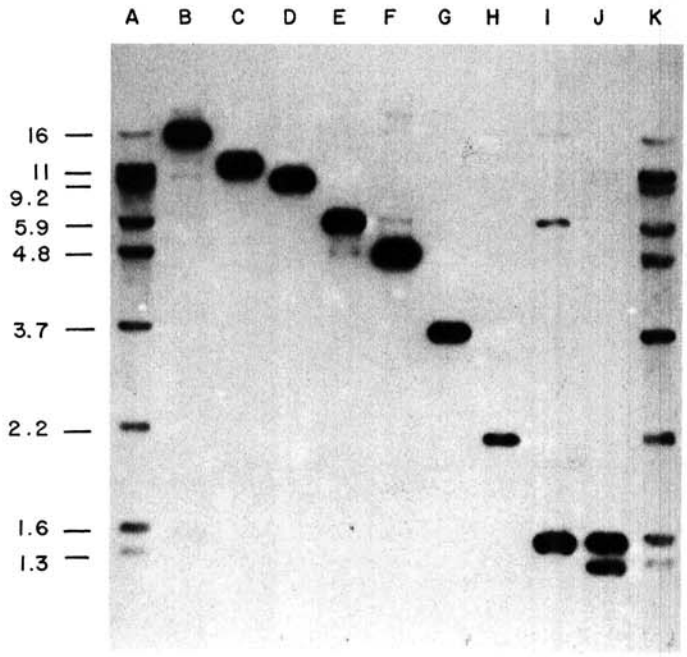

Fig. 4. Dispersal of repeated sequences in the genome of strain USDA257. Cosmid clones containing homology to the 2.3-kb *Hin*dIII/*Eco*RI insert of pHBK312 were identified by colony hybridization. DNA from individual clones was digested with *Eco*RI, electrophoresed, and blotted to nitrocellulose. The membrane was probed with the 0.95-kb *Hin*dIII/*Sal*I insert of pHBK311. Lanes A and K, genomic DNA of USDA257; lane B, pRFRS1; lane C, pRFRS2; lane D, pRFRS3; lane E, pRFRS4; lane F, pRFRS5; lane G, pRFRS6; lane H, pRFRS7; lane I, pRFRS8; lane J, pRFRS9. Sizes in kilobases.

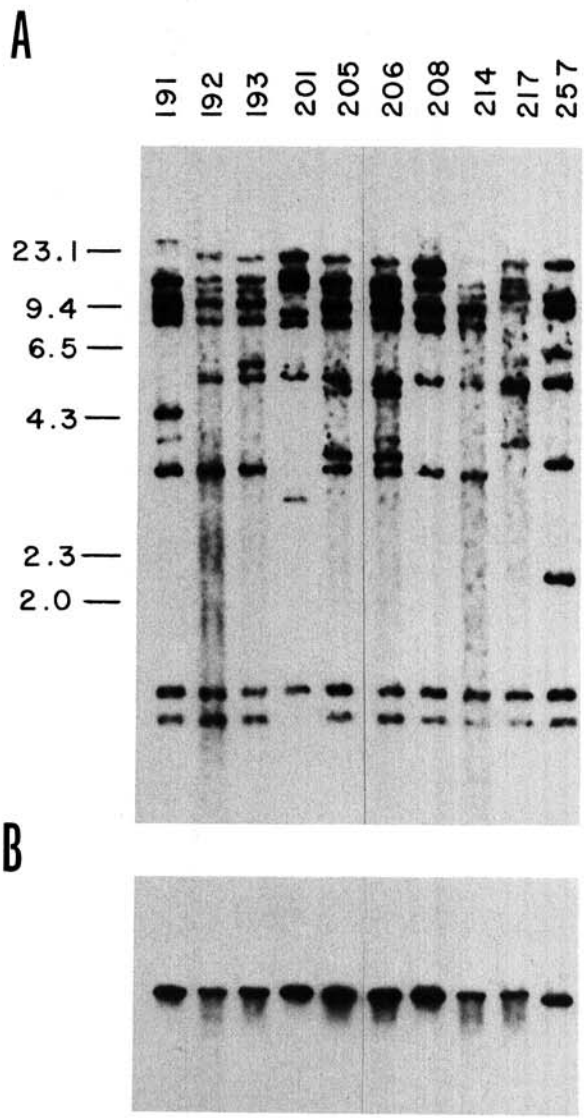

Fig. 5. Distribution of repetitive sequences in *Rhizobium fredii*. Genomic DNA from the indicated strains (all prefixed USDA) was digested with *Eco*RI. Blots were probed with **A**, the 0.95-kb *Hin*dIII/*Sal*I fragment, and **B**, the 1.7-kb *Bam*HI fragment that contains *nodABC* from USDA257. Sizes in kilobases.

pHBK303 (Fig. 7) in the region known to be homologous to multiple bands within the USDA257 genome. As shown in Figure 1, this sequence began at the left *Eco*RI border of the pHBK303 insert and extended to just beyond the *Hin*dIII site upstream from *nodA*. It contained two large overlapping ORFs on opposite strands. ORF1 began at position 842 and terminated at position 1913; the deduced protein was hydrophilic and 40.6 kD in size. ORF2 began at position 1771 and terminated at position 790; the deduced protein was hydrophilic and 34.7 kD in size.

Both ORFs were within a large, centrally located region (marked with bent arrows in Fig. 7) that has strong homology to the T-DNA of *A. rhizogenes* (Slightom *et al.* 1986). The limits of this region were defined by ORF1 and ORF2. Thus, homology began at position 772, only 16 bp beyond the termination codon for ORF2, and it ended precisely with the termination codon for ORF1 at position 1913. The relatedness of the sequences from the two species was considerable within this entire region and included conserved stretches as large as 35 bp in length. Putative genes homologous to both ORFs from *R. fredii* were present in the *A. rhizogenes* sequence. ORF1 of *R. fredii* corresponds to ORF6 of *A. rhizogenes* (Slightom *et al.* 1986) and has 74% sequence homology to it. The homology between the deduced proteins was similarly striking: 74% of the amino acid sequences were identical and 81% shared charge similarity (Fig. 8). This relationship was greatest in the central and C-termini, where several 20–30 amino acid stretches were faithfully conserved.

ORF2 of *R. fredii* encompasses ORF7 of *A. rhizogenes* (Slightom *et al.* 1986) and has 73% sequence identity to it. The ORF2 protein, however, was potentially much larger than the ORF7 protein. The methionine at nucleotide position 1342 (Fig. 7) was encoded by both sequences, but multiple termination codons upstream from this site did not allow transcription of ORF7 to begin any earlier. *R. fredii* lacked the corresponding termination signals, and the N-terminal of the putative ORF2 transcript thus was extended by 143 amino acids.

The sequenced region from USDA257 also has very strong homology to a family of closely related repetitive sequences from *B. japonicum* (Kaluza *et al.* 1985). Nucleotide homology began at position 842, the initiation codon for ORF1, and extended through position 849 (Fig. 7). Within this region, the *R. fredii* sequence had greatest homology, 57%, with the *B. japonicum* sequence termed RSRjα10. Not surprisingly, RSRjα10 also had a substantial 56% homology with the corresponding *A. rhizogenes* sequences. In *B. japonicum*, RSRjα10 is flanked by inverted repeats and contains a single ORF (ORF355) that is likely to be a coding region (Kaluza *et al.* 1985; Ramseier and Göttfert, in press). Although the inverted repeats do not lie within the region of homology with *R. fredii*, most of the sequences corresponding to ORF355 are present.

Perhaps the most striking characteristic of the sequenced region from *R. fredii* was the presence of numerous direct and inverted repeats. There were seven different 10- and 11-bp direct repeats, as well as 58 different 8- to 11-bp inverted repeats within the 2,432-bp sequence. Included in the latter group were five palindromic sequences, one of which encompassed the *Hin*dIII site. Table 2 summarizes the sequences and locations of all of the 10- and 11-bp repeats. These repeats, as well as shorter repeats not included in the table, were scattered throughout the sequenced region. Although the positions of the repeats bear no apparent spatial relationship to one another, many appeared on both sides of the adjacent *Sal*I sites (Fig. 1).

Neither the RSRjα10 homologous region nor the T-DNA homologous region was flanked by duplications or pairs of direct or inverted repeats. There were nevertheless overlapping inverted repeats adjacent to both borders of the T-DNA homologous region:

```
758                                          1,944
 |                                             |
  GCCGTCGGCGCTGAAGC-- T-DNA homology
--TAGTAAGGGGTGGCTGAGGAACTTGCCGGCGT
```

On the left side, the overlapping 8-mers beginning with CGT and GGC were repeated at positions 1,686 and 1,203, respectively. On the right side, the sequences beginning with ACT, CTT, and CGT were repeated at positions 999, 44, and 2,050, respectively. None of the sequences, however, bears close structural resemblance to target sites for insertion sequencelike elements (Iida *et al.* 1983).

Northern analysis. Northern analysis was performed to test the hypothesis that ORF1 and OFR2 of *R. fredii* are transferred to the plant genome during nodulation. Total,

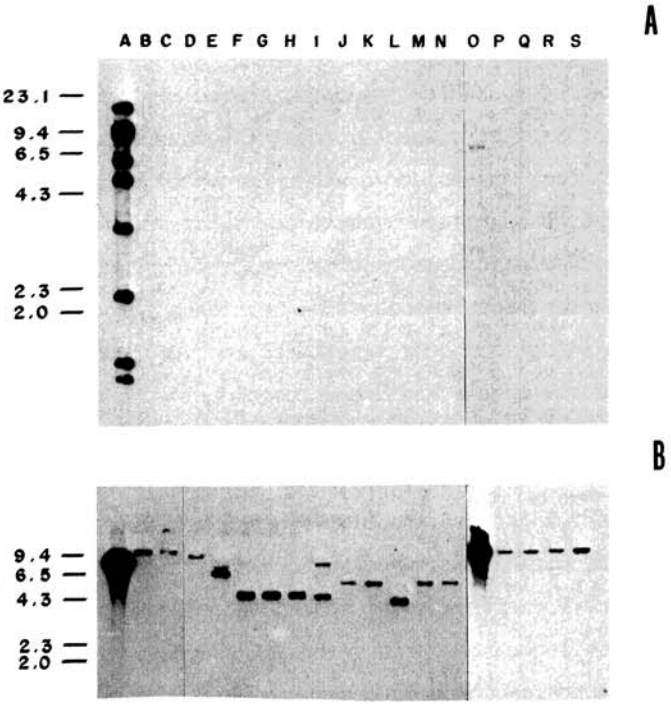

Fig. 6. Search for repetitive sequences in fast-growing rhizobia. Genomic DNA was digested with *Eco*RI and blots were probed with the 0.95-kb *Hin*dIII/*Sal*I fragment (**A**) or the 1.7-kb, *nodABC*-containing *Bam*HI fragment (**B**) from USDA257. Strains are as follows: (A) USDA257; (B) *Rhizobium leguminosarum* bv. *trifolii* USDA2050; (C) USDA2227; (D) BAL; (E) *R. leguminosarum* bv. *phaseoli* USDA2676; (F) USDA2667; (G) USDA2720; (H) USDA2693; (I) USDA2717; (J) *R. leguminosarum* bv. *viciae* 3HOq1; (K) PF_2; (L) PRE; (M) ANKK319; (N) TOM; (O) *Rhizobium* sp. NGR234; (P) *R. meliloti* 102F29; (Q) 102F51; (R) RCR2011; (S) 102F65. Sizes in kilobases.

poly(A)$^+$, and poly(A)$^-$ RNA from nodules was probed with the 0.95-kb *Hin*dIII/*Sal*I fragment (for ORF1 and ORF2) as well as pLB23, a positive control probe for leghemoglobin (Fuller and Verma 1984) (Fig. 9). pLB23 hybridized to a single, approximately 675-base transcript present only in total nodule RNA and the poly(A)$^+$ fraction derived from it. The probe for ORF1 and ORF2 hybridized to a series of relatively long transcripts. These transcripts were present only in nodules and were absent in the poly(A)$^+$ fraction.

DISCUSSION

A number of intriguing possibilities are raised by the discovery of reiterated DNA sequences and ORFs homologous to T-DNA in the immediate vicinity of *nodABC*. Foremost among these are questions of function: Do the reiterated sequences promote recombination or otherwise regulate nodulation? Or is their location merely accidental? DNA reiterations are in fact relatively common features of the genomes of rhizobia, where they encompass three broad categories. The first of these includes straightforward duplication of entire genes (e.g., *nodD* or *nifH*) (Young and Johnston 1989), or in some cases, just promoter regions (Better *et al.* 1983; Watson and Schofield 1985; Spaink *et al.* 1987). Insertion elements and related structures, including IS*Rm1*, IS*Rm2*, and IS*Rm3* (Ruvkun *et al.* 1982; Dusha *et al.* 1987; Wheatcroft and Laberge 1991) and the RSRjα family (Kaluza *et al.* 1985), make up the second group (Labes and Simon 1990; Simon *et al.* 1991). The last group includes duplications of unknown structure and function, most of which have been identified simply on the basis of homology among restriction fragments (Masterson and Atherly 1986; Flores *et al.* 1987).

The DNA reiterations described here were accidentally discovered when we probed the genome of *R. fredii* USDA257

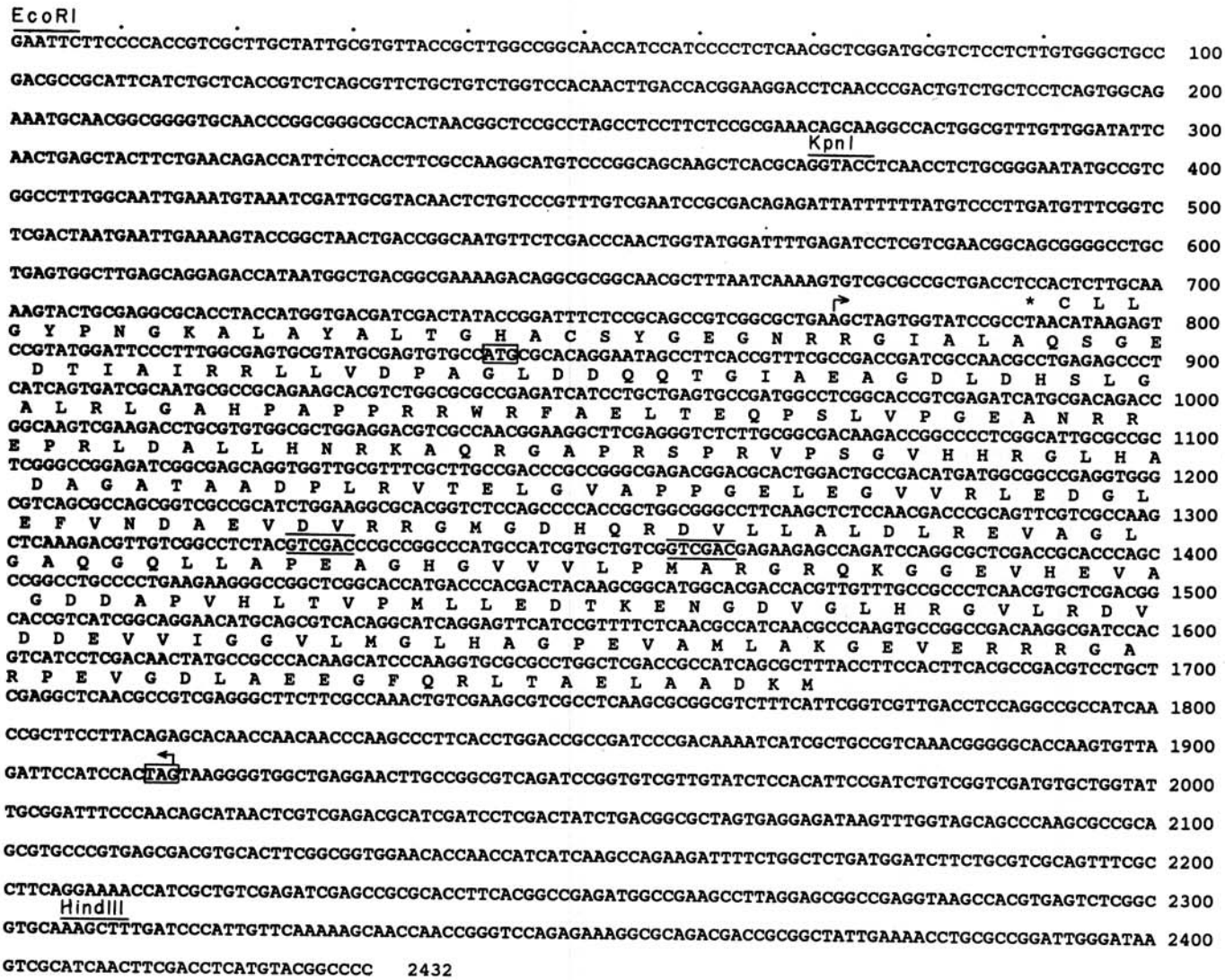

Fig. 7. Sequence of the *nodABC*-linked fragment that contains repetitive sequences. The region covers 2,432 bp, beginning with the *Eco*RI site at position 1. Other key restriction sites are labeled, or in the case of *Sal*I, encased in hortizontal lines. Arrows linked to vertical bars delineate the extent of homology to the T-DNA of *Agrobacterium rhizogenes*. The initiation and termination codons for ORF1 are boxed, and the entire translation for ORF2 is given.

with a restriction fragment containing *nodABC*. *nodABC* itself is not duplicated, and we could identify no diagnostic structural features that might represent flanking sequences for insertion elements (Iida *et al.* 1983) or promoters for known *nod* or *nif* genes. Rather, there were numerous short, repetitive sequences upstream from *nodA*. These reiterations were variable and complex, and although similar sets of sequences have not been described previously in rhizobia, our hybridization data indicated that they may be ubiquitous in *R. fredii*.

We do not yet understand the role of the short, repeated sequences within the genome of *R. fredii*. Initial analysis of a set of cosmids containing homology to the repeated sequences suggested genetic linkage, both between regions and to *nif* genes. This implied that multiple copies of the repetitive sequences were on the sym plasmid. One possibility is that these copies function to promote genetic rearrangements that influence nodulation, as is the case in other rhizobia (Ruvkun *et al.* 1982; Flores *et al.* 1987; Hahn and Hennecke 1987a; Soberón-Chávez and Nájera 1989; Brom *et al.* 1991). We have a series of genetically aberrant mutants of USDA257 that are derived from a Tn*5* mutagenesis program, and several had symbiotic abnormalities that cannot be traced to insertion of the transposon (Heron *et al.* 1989). Although this could be explained by secondary genetic rearrangements involving repetitive sequences, our hybridization data have not yet provided evidence to support this possibility.

One also can ask why a subset of T-DNA sequences, corresponding almost precisely to a pair of partially overlapping ORFs, is present in an organism not known to transfer DNA to plants. The *A. rhizogenes* reading frames, ORF6 and ORF7 (Slightom *et al.* 1986), are in a region of Ri T-DNA known to be expressed in transformed plants (Ooms *et al.* 1985; Taylor *et al.* 1985). The presence of this region, however, is unnecessary for transformation to occur (Durand-Tardif *et al.* 1985). ORF6 and ORF7 are structurally distinct. ORF6 is associated with eukaryotic ribosome-binding sites and promoters, as well as putative polyadenylation signals; ORF7 lacks these eukaryotic features and more closely resembles a prokaryotic gene. These features have led Slightom *et al.* (1986) to propose that ORF6 is transcribed in plants and that ORF7 may be transcribed in *A. rhizogenes*, perhaps as a regulatory element.

ORF1 and ORF2 of *R. fredii* USDA257 share strong sequence homology with ORF6 and ORF7, respectively. We could find no evidence for the existence in nodules of poly(A)$^+$ transcripts corresponding to either of the *R.*

```
F-RS   1 MRTGIAFTVSPTDRQRLRALISDRNAPQKHVWRAEIILLSADGLGTVEIM
         |||||.|.|.:.||.|| |::.....:|||||||.|||:|.|| |||.||
R-Ti   1 MRTGITFNVAAADRARLEAIVAAPTSAQKHVWRAKIILMSSDGSGTVAIM

        51 RQTGKSKTCVWRWRTS.PTEGFEGLLRRQDRPLGIAPLGPEIGEQVVAFR
             .||||||||||||...  |||.:|||: ..|| |||||:.|:.|.|||:
        51 EATGKSKTCVWRWQERFMTEGVDGLLHDKSRPPGIAPLDGELVERVVALT

       100 LPTRRARRTHWTADMMAAEVGVSASG.RRIWKAHGLQPHRWRAFKLSNDP
           |.|...  ||||.  || .||:.||:  :||..|||.|||||.|||||.
       101 LETPQQEATHWTVRAMAKAVGIAASSVVKIWHEHGLAPHRWRSFKLSNDK

       149 QFVAKLKDVVGLYVDPPAHAIVLSVDEKSQIQALDRTQPGLPLKKGRLGT
           .|..||.|||||||.|||||||||||||||||||||||||||||||||||
       151 AFAEKLHDVVGLYVSPPAHAIVLSVDEKSQIQALDRTQPGLPLKKGRAGT

       199 MTHDYKRHGTTTLFAALNVLDGTVIGRNMQRHRHQEFIRFLNAINAQVPA
           |||||||||||||||||:|||.|||||||||||||||||||:|::|
       201 MTHDYKRHGTTTLFAALNILDGSVIGRNMQRHRHQEFIRFLNAIEAELPK

       249 DKAIHVILDNYAAHKHPKVRAWLDRHQRFTFHFTPTSCSRLNAVEGFFAK
           |||:|||||||.|:|||||||.|.|:|||||.|||||:|||||||||||
       251 DKAVHVILDNYATHKQPKVRAWLARHPRWTFHFVPTSCSWLNAVEGFFAK

       299 LSKRRLKRGVFHSVVDLQAAINRFLTEHNQQPKPFTWTADPDKIIAAVKR
           |.:||||:|||||||||||.|||||: ||||:|||||.| ||||.|||||||
       301 LTRRRLKHGVFHSVVDLQATINRFVREHNQEPKPFIWRADPDEIIAAVKR

       349 GHQVLDSIH
           |||.|:|||
       351 GHQALESIH
```

Fig. 8. Comparison of the deduced amino acid sequences of ORF1 from USDA257 (designated RF-RS) and ORF6 from the T-DNA of *Agrobacterium rhizogenes* (designated AR-Ti). Vertical bars indicate identity. Single and double dots respectively indicate similarity and high similarity of corresponding amino acids. The *A. rhizogenes* sequence is from Slightom *et al.* (1986).

Table 2. Ten and 11-bp direct and inverted repeats in the 2,432-bp repetitive sequence containing region of USDA257

Type	Size (bp)	Sequence	Positions
Direct	11	CCTGCTGAGTG	596, 952
Direct	11	CTGTCGGTCGA	1,352, 1,979
Direct	11	CGACCCGCCGG	1,143, 1,327
Direct	11	GCGGCCGAGGT	1,187, 2,271
Inverted	11	TGATGGCGGCC	1,182, 1,787
Direct	10	CGGCACCGTC	973, 1,498
Direct	10	GCTCGACCGC	1,384, 1,651
Direct	10	CGCCATCAAC	1,561, 1,792
Inverted	10	CAAAGCTTTG	2,304[a]
Inverted	10	AAGGTGCGCG	1,637, 2,233
Inverted	10	CGCCAGCGGT	1,207, 1,251
Inverted	10	GATGCGTCTC	76, 2,029

[a] Palindromic sequence.

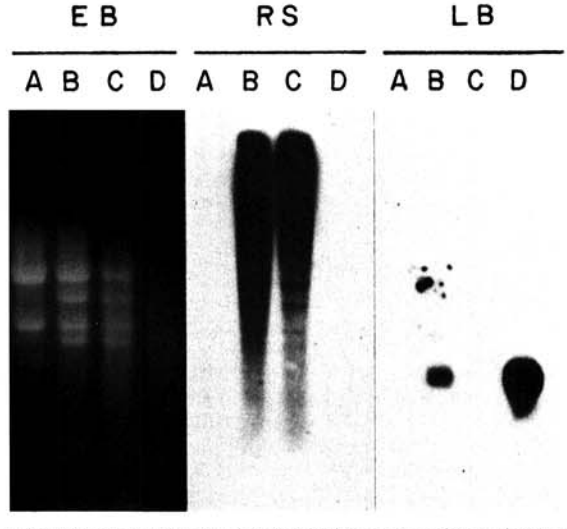

Fig. 9. Northern analysis showing the ethidium bromide-stained gel (EB) and filters probed with the 0.95-kb *Hin*dIII/*Sal*I fragment from USDA257 (RS) and a leghemoglobin probe (LB). **A,** total root RNA; **B,** total nodule RNA; **C,** poly(A)$^-$ nodule RNA; **D,** poly(A)$^+$ nodule RNA. Lanes A–C are loaded with 14 μg of RNA; lane D contains 3 μg of RNA. The bar marks the position of the 675 base leghemoglobin transcript.

fredii ORFs. This observation is fully consistent with the absence of eukaryoticlike promoters and polyadenylation signals in association with the ORFs from *R. fredii*, and it rules out the possibility that ORF1 and ORF2 represent genes that are transferred to the plant during nodulation. Thus, if these ORFs have any function, it must be in the bacterium itself.

R. fredii USDA257 is not the only nitrogen-fixing organism with homology to ORF6 and ORF7 of *A. rhizogenes*. Hennecke and associates have described a family of repetitive sequences, termed RSα, in *B. japonicum* (Kaluza *et al.* 1985; Hahn and Hennecke 1987a, 1987b). Related sequences exist in a variety of rhizobia, including *R. fredii* USDA191 (Hahn and Hennecke 1987a). The hybridization pattern produced when *Eco*RI-digested genomic DNA of USDA191 is probed with RSα is nearly identical to that obtained with the 0.96-kb *Hin*dIII/*Sal*I probe from USDA257 (cf. Fig. 5 with Fig. 3 of Hahn and Hennecke 1987a). Alignment of our sequence with those of the RSα family members readily established the basis for the observed *B. japonicum*/*A. rhizogenes* homology. Computer analysis additionally substantiates the considerable homology between T-DNA and RSα, a relationship that was initially overlooked (Kaluza *et al.* 1985) but has been more recently recognized (Ramseier and Göttfert, in press). The RSRjα sequence contains a single ORF that is likely to be transcribed (Ramseier and Göttfert, in press). This ORF is well conserved in both *A. rhizogenes* and *R. fredii*, but the insertion elementlike direct repeats that flank RSα are present only in *A. rhizogenes* (Kaluza *et al.* 1985; Slightom *et al.* 1986; Ramseier and Göttfert, in press).

The above data suggest that ORF1 and ORF2 may represent primordial genes that, for unknown reasons, are conserved in several divergent members of the Rhizobiaceae. Such primitive genes lack internal redundancies (Levesque *et al.* 1988) and could theoretically be either prokaryotic or eukaryotic in origin. Although the Ri T-DNA homology raises the potential of eukaryotic antecedents, the *R. fredii* sequences lack eukaryotic features and seem to be prokaryotic.

ACKNOWLEDGMENTS

This is Journal Series paper 11418 of the Missouri Agricultural Experiment Station and is supported by USDA Grant 90-37262-5780 and by the Food for the 21st Century Program, University of Missouri.

LITERATURE CITED

Aviv H., and Leder, P. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. Proc. Natl. Acad. Sci. USA 69:1408-1412.

Balatti P., and Pueppke, S. G. 1990. Nodulation of soybean by a transposon-mutant of *Rhizobium fredii* is subject to competitive nodulation blocking by other rhizobia. Plant Physiol. 94:1276-1281.

Better, M., Lewis, B., Corbin, D., Ditta, G., and Helinski, D. R. 1983. Structural relationships among *Rhizobium meliloti* symbiotic promoters. Cell 35:479-485.

Brom, S., Garcia de los Santos, A., de Lourdes Girard, M., Dávila, G., Palacios, R., and Romero, D. 1991. High frequency rearrangements in *Rhizobium leguminosarum* bv. *phaseoli* plasmids. J. Bacteriol. 173:1344-1346.

Cannon, F. C., Riedel, G. E., and Ausubel, F. M. 1979. Overlapping sequences of *Klebsiella pneumoniae nif* DNA cloned and characterized. Mol. Gen. Genet. 174:59-66.

Durand-Tardif, M., Broglie, R., Slightom, J., and Tepfer, D. 1985. Structure and expression of Ri T-DNA from *Agrobacterium rhizogenes* and *Nicotiana tabacum*. J. Mol. Biol. 186:557-564.

Dusha, I., Kovalenko, S., Banfalvi, Z., and Kondorosi, A. 1987. *Rhizobium meliloti* insertion element IS*Rm2* and its use for identification of the *fixX* gene. J. Bacteriol. 169:1403-1409.

Flores, M., González, V., Brom, S., Martínez, E., Piñero, D., Romero, D., Dávila, G., and Palacios, R. 1987. Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol. 169:5782-5788.

Fuller, F., and Verma, D. P. S. 1984. Appearance and accumulation of nodulin mRNAs and their relationship to the effectiveness of root nodules. Plant Mol. Biol. 3:21-28.

Hahn, M., and Hennecke, H. 1987a. Conservation of a symbiotic DNA region in soybean root nodule bacteria. Appl. Environ. Microbiol. 53:2253-2255.

Hahn, M., and Hennecke, H. 1987b. Mapping of a *Bradyrhizobium japonicum* DNA region carrying genes for symbiosis and an asymmetric accumulation of reiterated sequences. Appl. Environ. Microbiol. 53:2247-2252.

Heron, D. S., and Pueppke, S. G. 1984. Mode of infection, nodulation specificity, and indigenous plasmids of 11 fast-growing *Rhizobium japonicum* strains. J. Bacteriol. 160:1061-1066.

Heron, D. S., Érsek, T., Krishnan, H. B., and Pueppke, S. G. 1989. Nodulation mutants of *Rhizobium fredii* USDA257. Mol. Plant-Microbe Interact. 2:4-10.

Iida, S., Meyer, J., and Arber, W. 1983. Prokaryotic IS elements. Pages 159-221 in: Mobile Genetic Elements. J. A. Shapiro, ed. Academic Press, New York.

Jagadish, M. N., and Szalay, A. A. 1984. Directed transposon Tn*5* mutagenesis and complementation in slow-growing, broad host range cowpea *Rhizobium*. Mol. Gen. Genet. 196:290-300.

Kaluza, K., Hahn, M., and Hennecke, H. 1985. Repeated sequences similar to insertion elements clustered around the *nif* region of the *Rhizobium japonicum* genome. J. Bacteriol. 162:535-542.

Keyser, H. H., Bohlool, B. B., Hu, T. S., and Weber, D. F. 1982. Fast-growing rhizobia isolated from root nodules of soybean. Science 215:1631-1632.

Krishnan, H. B., and Pueppke, S. G. 1991. Sequence and functional analysis of the *nodABC* region of *Rhizobium fredii* USDA257, a broad host range nitrogen-fixing symbiont of soybean and other legumes. Mol. Plant-Microbe Interact. 5:512-520.

Labes, G., and Simon, R. 1990. Isolation of DNA insertion elements from *Rhizobium meliloti* which are able to promote transcription of adjacent genes. Plasmid 24:235-239.

Levesque, H., Delepelaire, P., Rouze, P., Slightom, J., and Tepfer, D. 1988. Common evolutionary origin of the central portions of the Ri T-DNA of *Agrobacterium tumefaciens*. Plant Mol. Biol. 11:731-744.

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

Martinez, E., Romero, D., and Palacios, R. 1990. The *Rhizobium* genome. Critic. Rev. Plant Sci. 9:59-93.

Masterson, R. V., and Atherly, A. G. 1986. The presence of repeated DNA sequences and a partial restriction map of the pSym of *Rhizobium fredii* USDA193. Plasmid 16:37-44.

Ooms, G., Twell, D., Bossen, M. E., Hoge, J. H. C., and Burrell, M. M. 1986. Developmental regulation of RI T-DNA gene expression in roots, shoots and tubers of transformed potato (*Solanum tuberosum* cv. Desiree). Plant Mol. Biol. 6:321-330.

Prakash, R. K., and Atherly, A. G. 1984. Reiteration of genes involved in symbiotic nitrogen fixation by fast-growing *Rhizobium japonicum*. J. Bacteriol. 160:785-787.

Priefer, U. B., Burkardt, H. J., Klipp, W., and Pühler, A. 1981. IS*RI*: An insertion element isolated from the soil bacterium *Rhizobium lupini*. Cold Spring Harbor, Symp. Quant. Biol. 45:87-91.

Quinto, C., de la Vega, H., Flores, M., Fernández, L., Ballado, T., Soberón, G., and Palacios, R. 1982. Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. Nature (London) 299:724-726.

Quinto, C., de la Vega, H., Flores, M., Leemans, J., Cevallos, M. A., Pardo, M. A., Azpiroz, R., Girard, M. D. L., Calva, E., and Palacios, R. 1985. Nitrogenase reductase: A functional multigene family in *Rhizobium phaseoli*. Proc. Natl. Acad. Sci. USA 82:1170-1174.

Ramseier, T. E., and Göttfert, M. Codon usage and G + C content in *Bradyrhizobium japonicum* genes are not uniform. Arch. Microbiol.

In press.
Rochester, D. E., Winter, J. A., and Shah, D. M. 1986. The structure and expression of maize genes encoding the major heat shock protein hsp 70. EMBO J. 5:451-458.
Ruvkun, G. B., Long, S. R., Meade, H. M., van den Bos, R. C., and Ausubel, F. M. 1982. IS*Rm1*: A *Rhizobium meliloti* insertion sequence that transposes preferentially into nitrogen fixation genes. J. Mol. Appl. Genet. 1:405-418.
Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
Scholla, M. H., Moorefield, J. A., and Elkan, G. H. 1984. Deoxyribonucleic acid homology between fast-growing soybean-nodulating bacteria and other rhizobia. Int. J. Syst. Bacteriol. 34:283-286.
Simon, R., and Priefer, U. B. 1989. Vector technology of relevance to nitrogen fixation research. Pages 13-49 in: The Molecular Biology of Symbiotic Nitrogen Fixation. P. H. Gresshoff, ed. CRC Press, Boca Raton, FL.
Simon, R., Hötte, B., Klauke, B., and Kosier, B. 1991. Isolation and characterization of insertion elements from Gram-negative bacteria by using new broad-host-range, positive selection vectors. J. Bacteriol. 173:1502-1508.
Slightom, J. L., Durand-Tardif, M., Jouanin, L., and Tepfer, D. 1986. Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid: Identification of open reading frames. J. Biol. Chem. 261:108-121.
Soberón-Chávez, G., and Nájera, R. 1989. Symbiotic rearrangement in a hyper-recombinant mutant of *Rhizobium leguminosarum* biovar *phaseoli*. J. Gen. Microbiol. 135:47-54.
Spaink, H. P., Okker, R. J. H., Wijffelman, C. A., Pees, E., and Lugtenberg, B. J. J. 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* sym plasmid pRL1JI. Plant Mol. Biol. 9:27-39.
Taylor, B. H., White, F. F., Nester, E. W., and Gordon, M. P. 1985. Transcription of *Agrobacterium rhizogenes* A4 T-DNA. Mol. Gen. Genet. 201:546-553.
Vincent, J. M. 1970. A Manual for the Practical Study of Root-Nodule Organisms. Blackwell Scientific Publications, Oxford.
Watson, J. M., and Schofield, P. R. 1985. Species-specific, symbiotic plasmid-located repeated DNA sequences in *Rhizobium trifolii*. Mol. Gen. Genet. 199:279-289.
Wheatcroft, R., and Laberge, S. 1991. Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence IS*Rm3*: Similarity between the putative transposase encoded by IS*Rm3* and those encoded by *Staphylococcus aureus* IS256 and *Thiobacillus ferrooxidans* IS*T2*. J. Bacteriol. 173:2530-2538.
Wheatcroft, R., and Watson, R. J. 1988. Distribution of insertion sequence IS*Rm1* in *Rhizobium meliloti* and other Gram-negative bacteria. J. Gen. Microbiol. 134:113-121.
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.