Organization and partial sequence of a DNA region of the *Rhizobium leguminosarum* symbiotic plasmid pRL6JI containing the genes *fix*ABC, *nif*A, *nif*B and a novel open reading frame

Petra Grönger, Sundaram S.Manian[1], Helmut Reiländer, Michael O'Connell[2], Ursula B.Priefer* and Alfred Pühler

Lehrstuhl für Genetik, Fakultät für Biologie, Universität Bielefeld, 4800 Bielefeld 1, FRG

Received October 21, 1986; Revised and Accepted December 8, 1986

ABSTRACT

By hybridization and heteroduplex studies the *fix*ABC and *nif*A genes of the *Rhizobium leguminosarum* symbiotic plasmid pRL6JI have been identified. DNA sequencing of the region containing *nif*A showed an open reading frame of 1557 bp encoding a protein of 56,178 D. Based on sequence homology, this ORF was confirmed to correspond to the *nif*A gene. Comparison of three *nif*A proteins (*Klebsiella pneumoniae*, *Rhizobium meliloti*, *Rhizobium leguminosarum*) revealed only a weak relationship in their N-terminal regions, whereas the C-terminal parts exhibited strong homology. Sequence analysis also showed that the *R.leguminosarum nif*A gene is followed by *nif*B and preceded by *fix*C with an open reading frame inserted in between. This novel ORF of 294 bp was found to be highly conserved also in *R.meliloti*. No known promoter and termination signals could be defined on the sequenced *R.leguminosarum* fragment.

INTRODUCTION

Biological nitrogen fixation, as carried out by a variety of prokaryotic organisms, is an intricate process regulated by complex genetic mechanisms. In *Klebsiella pneumoniae*, *nif*-specific repression is mediated by the *nif*L gene product (1,2) while the *nif*A protein is required for transcriptional activation of all *nif* operons, except its own, *nif*LA (3,4,5,6).

Within the last few years it has become obvious that the *nif*A gene plays a central regulatory role in symbiotic nitrogen fixing organisms also. In *Rhizobium meliloti*, a gene showing functional and sequence homology to the *K.pneumoniae nif*A gene has been identified and shown to be essential for the activation of *nif* and *fix* genes (7,8,9,10). Similar genes homologous to *K.pneumoniae* and *R.meliloti nif*A have also been identified in *Rhizobium trifolii* (13) and *Bradyrhizobium japonicum* (14). The *R.meliloti nif*A gene, also designated *fix*D, is located between

the *fix*ABC operon and the *nif*B gene and transcribed in the same orientation (8,9,10,11,12,15,16,17). The function of the *fix*ABC genes, which have also been localized in *Bradyrhizobium japonicum* (18), is still unknown. The *nif*B gene product has been reported to be essential for the synthesis of the FeMo-cofactor in *K.pneumoniae* (19,20) and probably, it plays a similar role in *Rhizobium*.

Genes homologous to *K.pneumoniae nif*A and *nif*B have also been identified in *R.leguminosarum*. In strain 248, a region of the symbiotic plasmid pRL1JI was shown to contain a *K.pneumoniae nif*B-like gene, called *fix*Z, as well as the last portion of an open reading frame homologous to the 3' end of *K.pneumoniae nif*A (21). For *R.leguminosarum* strain PRE, a *nif*A homologous gene has been identified on a 3.3 kb *Bam*HI fragment and shown to be necessary for the expression of the nitrogenase structural genes *nif*HDK. Hybridization to the *fix*Z region was observed next to the *nif*A homologous gene (22,23). In both strains of *R.leguminosarum*, a cluster of *fix* genes has been defined upstream of this region (22,23,24). Thus, a regulatory gene homologous to *K.pneumoniae nif*A appears to be present on the *R.leguminosarum* symbiotic plasmids pRL1JI and pPRE and to map, analogous to *R.meliloti*, between *nif*B and a cluster of *fix* genes.

In this study, a region of the *R.leguminosarum* symbiotic plasmid pRL6JI has been identified which hybridized to *R.meliloti fix*ABC, *nif*A and *nif*B and mapped approximately 10 kb away from the common *nod* genes. More substantial evidence for the presence of these *fix* and *nif* genes on pRL6JI was obtained by heteroduplex experiments and by DNA sequence comparison with the corresponding regions of *R.meliloti*.

Parts of this work including the entire amino acid sequence of the *R.leguminosarum nif*A protein were presented at the Third International Symposium on "The Molecular Genetics of Plant-Microbe Interaction" in Montreal, 1986.

MATERIAL AND METHODS

Strains and plasmids

The relevant characteristics of the bacterial strains and plasmids used in this study are listed in Table 1.

Table 1. Bacterial strains and plasmids used. All hybridization probes listed carry segments of the *R.meliloti* symbiotic plasmid.

	Relevant markers and characteristics	Source or Reference
Strains		
E.coli S17-1	Pro, Res, Mod, RP4-2(Tc::Mu)(Km::Tn7)	54
R.leguminosarum 3855	128C53 Sm, contains pRL6JI	53
B151	128C53 Cm, cured of pRL6JI	53
B164	128C53 Sm, contains pIJ1008	Brewin
R.meliloti 2011	wild type, Sm	Dénarié
Vectors		
pSUP202	pBR325 derived Ap, Cm, Tc, Mob	54
pSUP205	derivative of pSUP202, contains *cos*-site in Ap	54
pSVB20	sequencing vector, based on pUC8 (56)	Arnold
Plasmids		
PIJ1008	recombinant between pRL6JI and pVW5JI, Nod, Fix, Hup, Tra, Nm (Tn5)	53
Hybridization Probes		
pRmSL26	Tc (pLAFR1), *nod*DABC and adjacent regions	52
pRmR3	Tc (pACYC184), *fix*ABC and part of *nif*A	15
pRmW53	Tc (pACYC184), *nif*K	Weber/Pühler
pRmW54	Tc (pACYC177-Cm), *nif*A and most of *nif*B	10
pMA152	Tc (pACYC184), *nod*ABC and part of *nod*D	Aguilar/Pühler
pMA14.1	Ap (pACYC177), *nod*C	Aguilar/Pühler

Media and growth conditions

E.coli cells were grown in LB medium (25) at 37°C; *Rhizobium* cells were grown in TY medium (26) at 30°C. Antibotics were used at the following concentrations: for *E.coli*: tetracycline 5µg/ml,

chloramphenicol 50µg/ml, ampicillin 100µg/ml; for *Rhizobium*: tetracycline 5µg/ml, streptomycin 400µg/ml, neomycin 100µg/ml.

Standard techniques

Bacterial matings were carried out as described by Simon (27). Plasmid DNA was isolated either on large scale by Triton X-100 lysis and purified by CsCl-EtBr density gradient centrifugation or on small scale by rapid isolation procedures described previously (28). Total DNA from *Rhizobium* was isolated according to Meade et al. (29) with modifications (30). DNA cloning, restriction enzyme analyses and ligations were performed according to Maniatis et al. (25). Transformations were carried out as described earlier (31).

Construction and identification of a pIJ1008 cosmid bank

Plasmid pIJ1008 is a 300 kb recombinant derived from the symbiotic plasmid pRL6JI of strain 128C53 and the bacteriocin producing plasmid pVW5JI (53). Cosmid cloning was performed as follows: Plasmid pIJ1008 was transferred to *R.meliloti* 2011 and a gene bank of the resulting strain 2011(pIJ1008) was constructed in the cosmid vector pSUP205. Cosmids were then individually mobilized from *E.coli* strain S17-1 into *R.leguminosarum* strains carrying plasmid pRL6JI or pIJ1008. Transconjugants were screened for stable maintenance of the tetracycline resistance marker. The vector pSUP205 cannot replicate in *R.leguminosarum*; only those cosmids carrying segments of plasmid pIJ1008 were expected to be stably inherited as a result of cointegrate formation between the insert DNA and the plasmid present in the recipient. Hybridization experiments confirmed that the selected cosmids contained only segments of plasmid pIJ1008 and pRL6JI, respectively. *Eco*RI digested cosmid DNA hybridized only to total DNA from *R.leguminosarum* strains carrying either plasmid pRL6JI (strain 3855) or pIJ1008 (strain B164) but not to strain B151 cured of plasmid pRL6JI. An identical hybridization pattern was observed for *R.meliloti* strain 2011(pIJ1008). Under the same conditions, no hybridization occurred with total DNA from *R.meliloti* 2011. Hybridization studies also confirmed that the arrangement of the *Eco*RI fragments in the cosmid was identical to that found in plasmids pIJ1008 and pRL6JI.

Restriction mapping of clone Cos4

DNA of the cosmid clone Cos4 was digested with *Eco*RI and all fragments were individually subcloned into the vector pSUP202. Restriction maps of all the *Eco*RI fragments were generated for the enzymes *Bam*HI, *Bgl*II and *Xho*I and compared with the patterns obtained for Cos4. Ambiguities were resolved by hybridization. In addition, the partial restriction method of Buikema et al. (32) was used.

Southern blot hybridization

DNA fragments were separated on agarose gels and denatured with 0.5M NaOH/1.5M NaCl, rinsed with distilled water and neutralized with 3.0M NaCl/0.5 Tris pH7.0. DNA was then transferred to nitrocellulose filters as described by Southern (33). Plasmid or cosmid DNA was labelled with ^{32}P-CTP by nick translation reaction (34). All other steps were essentially as described by Jagadish and Szalay (35).

DNA electron microscopy

Heteroduplex molecules between *R.leguminosarum* and *R.meliloti* DNA were performed according to the procedure described by Burkardt and Pühler (36). Photographs were taken with a Hitachi HS9 electron microscope and molecules were measured using a graphic calculator.

DNA sequencing

Sequencing was performed according to the chemical degradation method (37). The 3.3 kb *Bam*HI fragment was cloned in both orientations into the sequencing vector pSVB20, constructed and kindly provided by W. Arnold (manuscript in preparation). Sequence data were obtained from shot gun fragments and defined restriction fragments cloned into various pSVB20 plasmid derivatives and from specific deletions constructed using a modified technique described by Hong (38). All restriction sites were confirmed by overlapping sequencing.

Computer analysis

DNA sequencing gels were read manually. Programs used for sequence database constructions and for nucleotide or amino acid sequence analyses were those described by Martinez (39) and Sobel and Martinez (40).

RESULTS

Identification and mapping of symbiotic genes on the *Rhizobium leguminosarum* plasmid pRL6JI

A partial cosmid bank of the *R.leguminosarum* plasmid pIJ1008, which is a recombinant plasmid derived from the symbiotic plasmid pRL6JI and the bacteriocin producing plasmid pVW5JI (53), was constructed and screened as described in Material and Methods. Initial hybridizations using *R.meliloti nod* (pRmSL26), *fix*ABC/*nif*A (pRmR3) and *nif*K (pRmW53) probes suggested that two of the cosmids (Cos49, Cos56) contained the *nif*HDK region, while one cosmid clone (Cos4) carried the *nod*DABC and *fix*ABC/*nif*A regions.

Having ascertained that Cos4 contained only sequences of the symbiotic plasmid pRL6JI which were colinear with the original arrangement, further experiments concentrated on a more accurate mapping of the genes located on this cosmid. An *Eco*RI digest of Cos4 was probed with pRmR3 containing the *R.meliloti fix*ABC/*nif*A region and with pRmW54 carrying the *R.meliloti nif*A/*nif*B genes. The hybridization patterns obtained, revealed that sequences homologous to the *R.meliloti fix*ABC/*nif*A probe resided on a 5.45 kb *Eco*RI fragment, while homology to *nif*A/*nif*B was found to both a 5.45 kb and a 1.80 kb *Eco*RI fragment (data not shown). Similarly, both *Eco*RI fragments hybridized to the 1.5 kb *Bam*HI fragment of pRmW54, which contains only *nif*A specific sequences. Hybridization experiments with plasmid pMA152 established that sequences homologous to the *R.meliloti nod*DABC region resided on two *Eco*RI fragments of sizes 2.95 kb and 1.35 kb; homology to *nod*C (probe pMA14.1) was only present on the 2.95 kb *Eco*RI fragment. The organization of Cos4 is shown in Figure 1. The *fix/nif* region of plasmid pRL6JI is approximately 10 kb from the *nod* region. Due to the absence of overlapping fragments between Cos4 and Cos49 or Cos56, it was not possible to determine the physical linkage between the *fix*ABC/*nif*A and the *nif*HDK regions.

Comparative analysis of the *R.leguminosarum* and *R.meliloti fix*ABC and *nif*A regions by heteroduplex studies

To compare the structural organization of the *fix*ABC/*nif*A containing regions in *R.leguminosarum* and *R.meliloti* heteroduplex molecules between the 5.45 kb *Eco*RI fragment of Cos4 and the 4.8

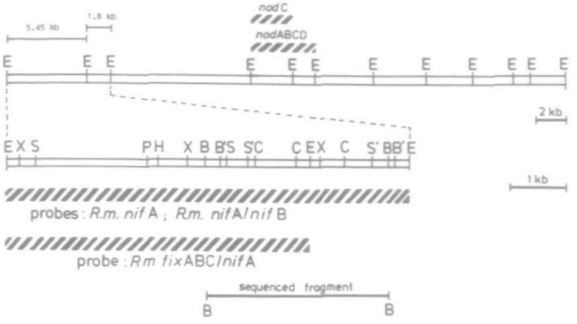


Figure 1. Organization of the pRL6JI cosmid clone Cos4. The restriction map for *Eco*RI is given at the top; the two *Eco*RI fragments (5.45 kb and 1.80 kb) hybridizing to *R.meliloti fix/nif* probes are shown below in more detail. Hatched areas refer to the sections hybridizing to the respective *R.meliloti* probes. The 3.3 kb *Bam*HI fragment used for sequencing is outlined. Restriction sites are B=*Bam*HI, B'=*Bgl*II, C=*Cla*I, E=*Eco*RI, H=*Hin*dIII, P=*Pst*I, S=*Sal*I, S'=*Sma*I, X=*Xho*I.

kb *Eco*RI fragment of pRmR3 were constructed. One of the 17 hybrid molecules analyzed is shown in Figure 2. The results indicated the presence of two regions of homology, a rather long section of approximately 3.5 kb and a very short one of 100-150 bp. Since the *R.meliloti* fragment was marked with Tn5, inserted into *fix*C, the long segment could be readily assigned to the coding region for *fix*ABC. However, the homology clearly extended beyond this region. The second segment of homology appeared within the *nif*A coding region but did not cover the entire fraction.

These data suggest that the organization of the *fix*ABC region is very similar in both *R.meliloti* and *R.leguminosarum* whereas the *nif*A genes are less conserved, at least in their 5' portions.

DNA sequence analysis of the *R.leguminosarum nif*A region

To better understand the dissimilarities within the *nif*A region of *R.leguminosarum* and that of *R.meliloti*, the 3.3 kb *Bam*HI fragment of Cos4, known to start within the region homologous to *fix*C and expected to contain the entire *nif*A gene of *R.leguminosarum* (Figs. 1 and 2) was sequenced. The cloning and sequencing strategy as well as the summary of the results are compiled in Figure 3. Figure 4 gives the complete nucleotide sequence ob-

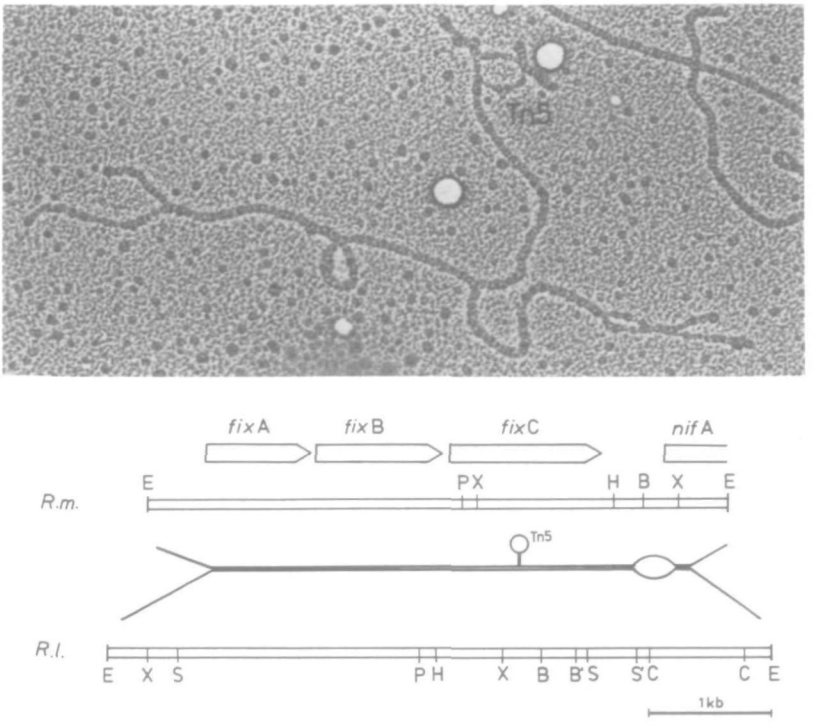


Figure 2. Heteroduplex prepared from the *fix*ABC/*nif*A regions of *R.leguminosarum* and *R.meliloti*. The electron micrograph shows the actual heteroduplex structure obtained between the 4.8 kb *Eco*RI fragment of pRmR3 (15) containing the *R.meliloti fix*ABC and part of the *nif*A gene and the 5.45 kb *Eco*RI fragment of Cos4 (compare also Fig. 1); the *R.meliloti* fragment was marked with transposon Tn5, inserted into *fix*C. The schematic representation shows the interpretation of the hybrid molecule, aligned to the maps of the respective *Eco*RI fragments. Restriction sites are as in Fig. 1.

tained for the 3.3 kb *Bam*HI fragment and the translation of open reading frames into amino acids. Four potential coding regions could be defined, all of which were in the same orientation.

One of these comprises 1557 nucleotides and contains a possible methionine initiation codon at position 989 and a termination codon at position 2545. The deduced amino acid sequence (519 amino acids) corresponds to a protein of 56,178 D. From the hybridization and heteroduplex studies it was very likely that this large open reading frame, overspanning the *Eco*RI restriction

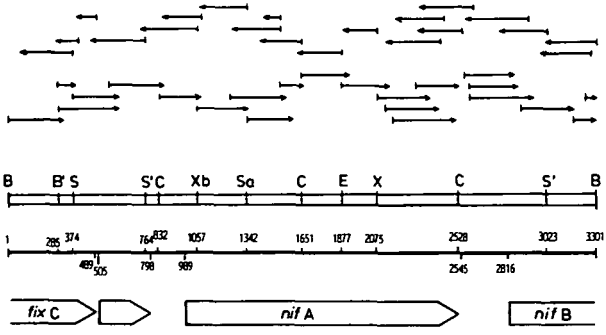
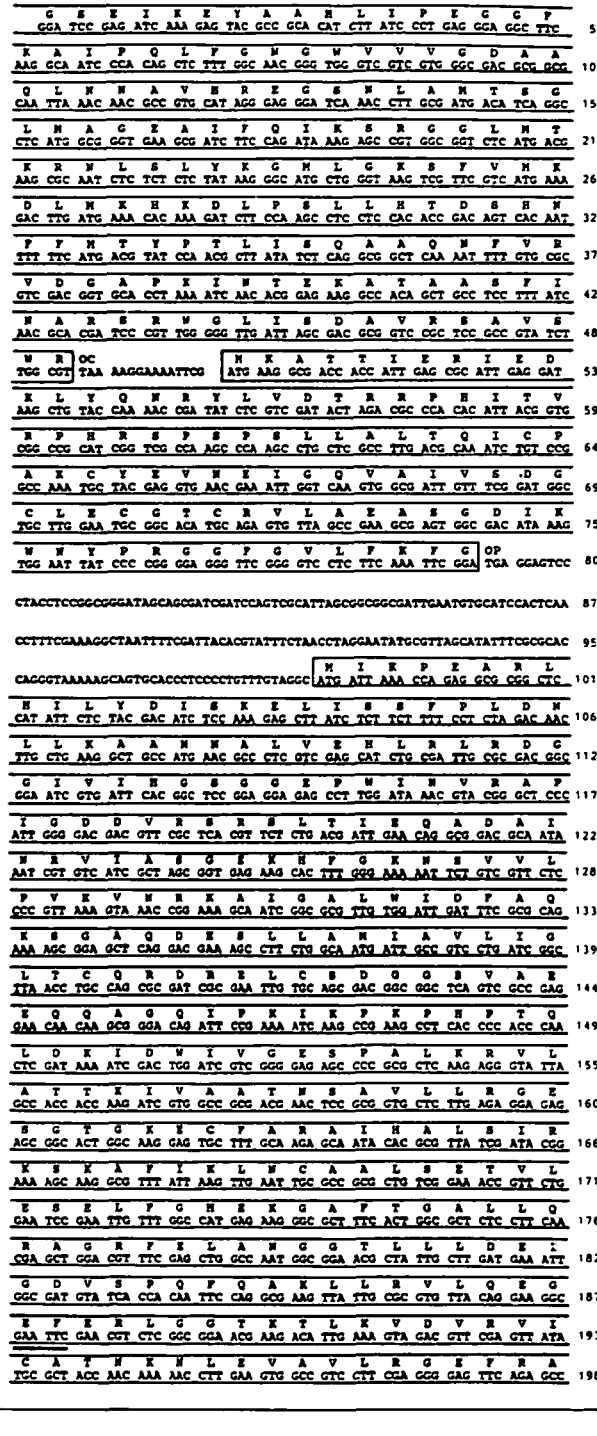


Figure 3. *Summary of the DNA sequencing of the pRL6JI fragment containing the nifA gene and adjacent regions. The 3.3 kb BamHI fragment of Cos4 (see also Fig. 1) was used for sequencing; arrows indicate the extent and direction of sequencing. In the middle part, the restriction map for the enzymes B=BamHI, B'=BglII, C=ClaI, E=EcoRI, S=SalI, S'=SmaI, X=XhoI, Xb=XbaI, Sa=SacI is outlined as obtained from the nucleotide sequence; the scale underneath refers to the relative positions of restriction sites (upper numbers) and of start and stop codons of open reading frames (bottom numbers). The location and transcriptional direction of open reading frames are shown below; three of them could be identified as fixC, nifA and nifB.*

site, corresponded to the *nif*A gene of *R.leguminosarum*. This was confirmed by aligning the predicted amino acid sequence with the *R.meliloti* (9,10) and *K.pneumoniae* (9,41) *nif*A sequences. Figure 5 presents this comparison with the sequences arranged in different domains, according to Drummond et al. (41). The different degrees of homology between the three *nif*A proteins for the various domains are summarized in Table 2 which also includes the comparison to the *K.pneumoniae ntr*C sequence (9,41). The first 140 amino acids of the *R.leguminosarum nif*A sequence (block A) show only weak homology with the other proteins, followed by a block of approximately 20 amino acids (domain C) with a slightly higher degree of homology. A central region of about 240 amino acids (domain D) exhibits strong correlation amongst the four proteins. One stretch of 16 amino acids (coordinates 285-300) is completely identical between the three *nif*A proteins and is also conserved in *ntr*C. Another section sharing substantial homology is near the end of the proteins (block E). The highest conservation to *R.meliloti* in this block occurs between position 537 and



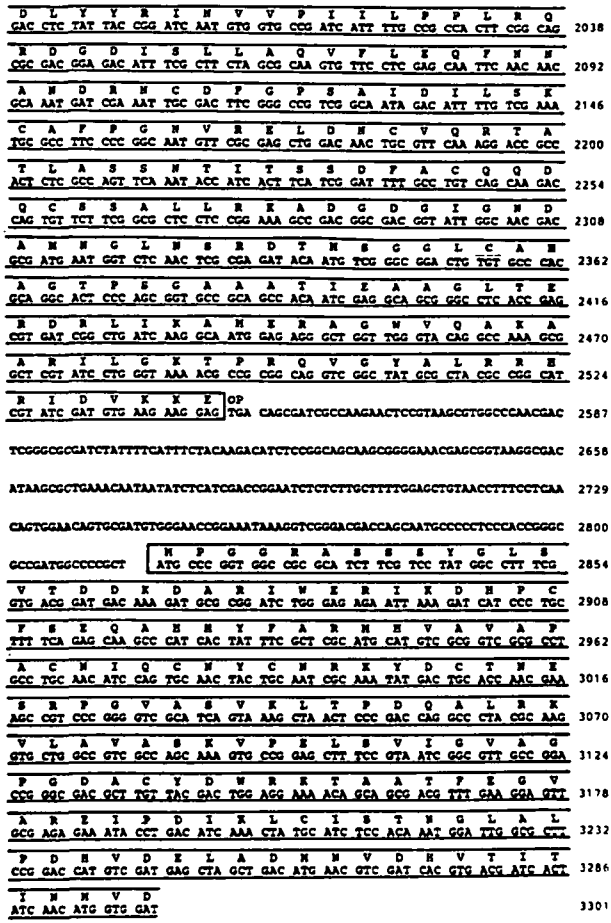


Figure 4. Complete sequence of the 3.3 kb *Bam*HI fragment containing *nif*A and adjacent regions. The 3.3 kb *Bam*HI fragment, known to start in a *fix*C homologous region and expected to contain the entire *nif*A gene was sequenced as described. Numbers refer to the nucleotide arrangement. The predicated amino acid sequences for the four open reading frames determined (boxed areas) are given in the single letter code. The unique *Eco*RI site located within the long open reading frame (position 1877) is underlined

562 with 24 out of 26 amino acids being identical.

From this comparison it was deduced that this long open reading frame corresponded to the *R.leguminosarum nif*A gene.

Localization of the *R.leguminosarum fix*C and *nif*B genes

Upstream of *nif*A, two open reading frames were localized,

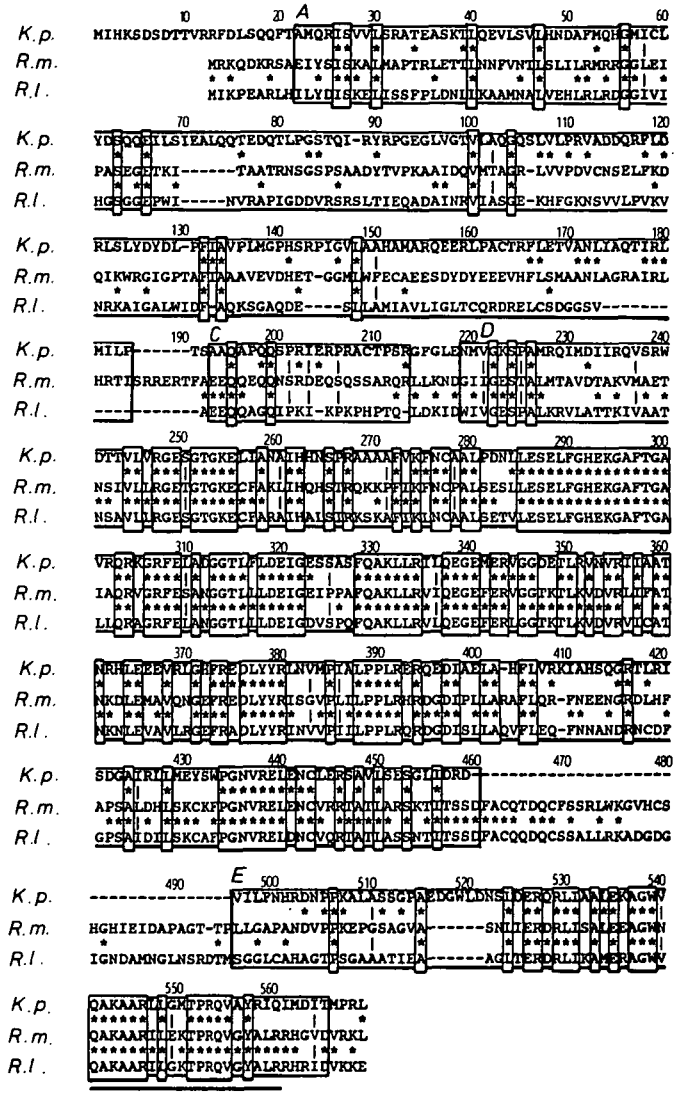


Figure 5. Comparison of the expected *nif*A amino acid sequence of *R.leguminosarum* (*R.l.*) with the *nif*A proteins of *R.meliloti* (*R.m.*, 9,10) and *K.pneumoniae* (*K.p.*, 9,41). Vertical bars denote identical amino acids in non-adjacent sequences, asterisks indicate identities in adjecent sequences. In addition, amino acids invariant in all three proteins are boxed. The alignment is according to Drummond et al. (41); the numbers above refer to the whole alignment, letters indicate blocks referred to in the text.

Table 2. Homology of the *R.leguminosarum nif*A amino acid sequence to *nif*A of *R.meliloti* (9,10) and *nif*A and *ntr*C of *K.pneumoniae* (9,41). The coordinates refer to the alignment in Figure 5. The percentage of identical residues was calculated for the various domains; n.d. = not determined since domain A is not present in *ntr*C (41).

Domain	Coordinates	Number of residues compared	Percentage of homology to *R.meliloti nif*A	*K.pneumoniae nif*A	*K.pneumoniae ntr*C
A	1-184	141	24.1	11.4	n.d.
C	193-213	19	31.6	26.3	15.8
D	219-460	240	72.9	55.8	44.6
E	496-565	64	59.4	46.9	21.5

one terminating at position 489 and a second very short one stretching between coordinates 505 and 798 (see Figs. 3 and 4). This small open reading frame (10,970 D) was very unlikely to correspond to *fix*C, since from minicell experiments, the *fix*C protein of *R.meliloti* was calculated to have a molecular weight of ca. 43,000 D (11). Therefore, the frame reading into the sequenced region was expected to represent the *R.leguminosarum fix*C coding region and was compared with the 3' end of the *R.meliloti fix*C gene, sequenced in our laboratory (H.R., unpublished results). The amino acid sequence alignment revealed a homology of 75.0% between the two species for the 44 amino acids compared (data not shown).

Downstream of the *nif*A gene, the N-terminal portion of another open reading frame was identified, starting at coordinate 2816. This open reading frame was expected to correspond to the beginning of the *nif*B locus, since it had been shown that in *R.leguminosarum*, as well as in *K.pneumoniae* and *R.meliloti*, the *nif*A gene is neighboured by a sequence homologous to *nif*B (21,23). The amino acids deduced from this sequence were therefore compared to the published sequence of *fix*Z, the *nif*B-like gene of *R.leguminosarum* plasmid pRL1JI (21). The alignment is shown in Figure 6. The sequence obtained in this work differs from that of Rossen et al. in that it starts 19 amino acids further upstream resulting in a different reading frame. Good correlation

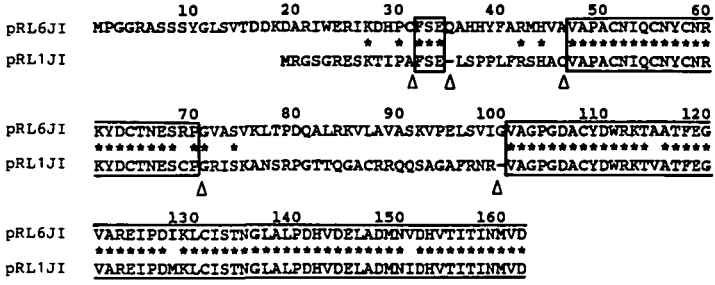


Figure 6. Alignment of the N-terminal amino acids of the *nif*B sequence (pRL6JI) to the *fix*Z sequence of pRL1JI (21). Asterisks denote identical residues; arrows indicate presumptive reading frame shift errors in the *fix*Z sequence. The sequences are arranged for the maximum matching of convergent reading frames (blocked areas).

between the two sequences begins at position 47 with another mismatch appearing between coordinates 70 and 100. However, the sequence obtained in this study is highly consistent with the *R.meliloti nif*B gene (H.R., manuscript in preparation), also in those sections differing from the *fix*Z sequence (124 amino acids being identical out of 162 residues compared = 76.5%). Putative frame shift errors in the *fix*Z sequence, as indicated in Figure 6, might be responsible for the divergency between the two *R.leguminosarum nif*B genes.

Identification of a novel open reading frame preceding the *nif*A genes in *R.leguminosarum* and *R.meliloti*

Unexpectedly, a fourth open reading frame of 294 nucleotides was found within the sequenced fragment, stretching between the *fix*C and the *nif*A gene. The heteroduplex experiment between the *fix*ABC/*nif*A regions of *R.leguminosarum* and *R.meliloti* (Fig. 2)

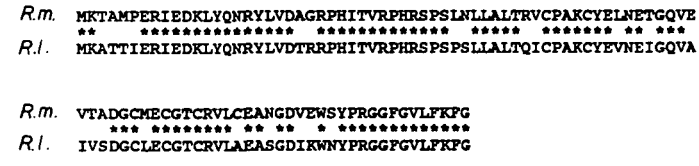


Figure 7. Comparison of the amino acid sequences of the novel open reading frames located between the *fix*C and *nif*A genes in *R.leguminosarum* (*R.l.*) and *R.meliloti* (*R.m.*). Asterisks indicate identical amino acids.

already suggested that this sequence might be conserved in the two species as the long stretch of homology corresponding to *fix*ABC clearly extended beyond the *fix*C coding regions. This assumption was confirmed by sequence analysis, which revealed the presence of an open reading frame downstream of the *fix*C gene of *R.meliloti*, which also comprises 98 amino acids, 77.5% of which are identical to those of *R.leguminosarum* (Fig. 7). In both species, this novel open reading frame is separated from the *fix*C coding region by only 12 nucleotides which include a good ribosome binding site (AAGGA in *R.leguminosarum*, GGAG in *R.meliloti*).

DISCUSSION

By hybridization to *R.meliloti* probes and confirmed by heteroduplex and sequencing studies, a fragment of the *R.leguminosarum* symbiotic plasmid pRL6JI containing the genes *fix*ABC, *nif*A and *nif*B and an additional open reading frame has been identified. The common *nod* gene cluster was mapped approximately 10 kb from the *fix/nif* region. This spacing is very similar to that found in the *R.leguminosarum* plasmid pPRE (22,23). In contrast, the intervening distance between the *nif*A homologue and the *nod*C gene in the *R.leguminosarum* plasmid pRL1JI is only 6 kb (24,55). Also in pRL5JI these regions are close together (42). Thus, symbiotic plasmids in *R.leguminosarum* seem to diverge with respect to their physical maps. In all *R.leguminosarum* symbiotic plasmids investigated, the *nif*A homologous gene was found to be flanked on one side by *fix*Z, the homologue of the *K.pneumoniae nif*B locus, and by a cluster of *fix* genes on the other side (21, 23,24).

The *R.leguminosarum nif*A protein (56,178 D) appears to be slightly smaller than that of both *R.meliloti* (59,868 D) (10) and *K.pneumoniae* (58,630 D) (41). Alignment of the *R.leguminosarum*, *R.meliloti* and *K.pneumoniae* amino acid sequences (according to Drummond et al., 41) shows a strong relationship between the three *nif*A proteins in their C-terminal part which is also significantly homologous to *ntr*C, although to a lower degree. The two *Rhizobium nif*A proteins are closer related to each other than to that of *K.pneumoniae*. The highest percentage of homology is found in a central region of approximately 240 residues (domain D),

followed by domain E which contains a sequence related to a DNA binding motif (underlined region in Fig. 5). Like in *K.pneumoniae nif*A, this binding motif in *R.leguminosarum* contains Gly at position 549, thus matching the consensus sequence proposed by Drummond et al. much better than that of *R.meliloti*. The homology in the N-terminal part of the three *nif*A proteins is less striking (domains A and C). The weak homology between *R.meliloti* and *K.pneumoniae nif*A in this area, already described by Drummond et al. (41), could be confirmed by including the comparison with *R.leguminosarum nif*A, the homology of which to *K.pneumoniae* is even lower.Although not necessarily expected, the two *Rhizobium nif*A proteins also exhibit only a weak relationship in their N-terminal parts, although the correlation is much better than that to *K.pneumoniae*. It would appear, that the highly conserved domains D and E are common to positive regulatory proteins, while the less conserved N-terminal domains may be responsible for species specific differences in their regulatory function.

No classical rho-independent terminator structures could be localized within the nontranslated regions upstream of *nif*A (187 bp) and *nif*B (267 bp). Further, no homology to *nif*A or *ntr*C regulated promoter sequences (12,43,44,45,46,47,48) or to the *R.meliloti nif*A promoter (9) was detected. Also, no sequences resembling the upstream elements shown to be required for *nif*A activation in some species could be identified (6,45,49,50). This is in contrast to *R.meliloti* where (i) *nif*A is preceded by a promoter sequence and can be expressed either from its own promoter or from the *fix*ABC promoter p2 (9,51) and (ii) the intergenic region between *nif*A and *nif*B contains both a rho-independent transcriptional termination signal and a *nif* consensus promoter (9, H.R., unpublished results). Also in the case of pRL1JI no evidence for the presence of a promoter in front of the *nif*B homologue *fix*Z could be found (21). These data suggest either, that in *R.leguminosarum* there are in fact no promoters in front of the *nif*A and *nif*B genes and transcription occurs exclusively from a promoter further upstream, or that there are promoters which show no homology to the *Rhizobium* promoters described so far.

In addition to *nif*A and parts of *fix*C and *nif*B, the sequenced fragment contained a novel open reading frame 5' to the

*nif*A gene. Hybridization to total cellular DNA revealed homologous regions also in a variety of other *R.leguminosarum* strains (data not shown). By sequence comparison, this small reading frame was found to be also present in *R.meliloti* and to be highly conserved at both nucleotide and amino acid levels. For the moment, there is no genetic evidence that this region is really transcribed or that the putative protein plays some role in nitrogen fixation. The close proximity to the *fix*C gene, the absence of transcriptional termination signals and promoter-like structures and the existence of good ribosome binding sites indicate only that this open reading frame constitutes a separate coding region, the expression of which is probably coupled to the *fix*ABC operon.

This study has shown that remarkable similarities exist between *R.leguminosarum* and *R.meliloti* in terms of the genetic and structural organization of the *fix*ABC, *nif*A and *nif*B regions. Dissimilarities as they occur in the intergenic regions or in the N-terminal part of the *nif*A genes may be responsible for species specific differences in regulation, expression and function of these genes.

ACKNOWLEDGEMENTS

We wish to thank W. Arnold for kindly providing the unpublished sequencing vectors and for help and advice in sequencing problems. We also want to thank B. Kosier and S. Malmivaara for their assistance in preparing the manuscript. M. O'Connell and H. Reiländer were supported by Agrigenetics Research Corporation, Boulder, Colorado; P. Grönger was financed by grant GBI-4-128-D of the European Community; S. S. Manian was the recipient of a training contract awarded by the CEC (GBI-146-D). Part of this work was supported by the BMFT.

Present addresses: [1]Department of Dairy and Food Microbiology, University College, Cork, Ireland; [2]National Institute for Higher Education, Dublin, Ireland

*To whom reprint requests should be sent

REFERENCES

1. Hill, S., Kennedy, C., Kavanagh, E., Goldberg, R. and Hanau, R. (1981) Nature **290**:424-426
2. Merrick, M., Hill, S., Hennecke, H., Hahn, M., Dixon, R. and Kennedy, C. (1982) Molec. Gen. Genet. **185**:75-81

3. Roberts, G.P. and Brill, W.J. (1980) J. Bact. **144**:210-216
4. Dixon, R., Eady, R., Espin, G., Hill, S., Iaccarino, M., Kahn, D. and Merrick, M. (1980) Nature **286**:128-132
5. Buchanan-Wollaston, V., Cannon, M.C., Beynon, J.L. and Cannon, F.C. (1981) Nature **294**:776-778
6. Drummond, M., Clements, J., Merrick, M. and Dixon, R. (1983) Nature **301**:302-307
7. Zimmerman, J.L., Szeto, W.W. and Ausubel, F.M. (1983) J.Bact. **156**:1025-1035
8. Szeto, W.W., Zimmermann, J.L., Sundaresan, V. and Ausubel, F.M. (1984) Cell **36**:1035-1043
9. Buikema, W.J., Szeto, W.W., Lemley, P.V., Orme-Johnson, W.H. and Ausubel, F.M. (1985) Nucleic Acids Research **13**:4539-4555
10. Weber, G., Reiländer, H. and Pühler, A. (1985) EMBO J. **4**: 2751-2756
11. Pühler, A., Aguilar, M.O., Hynes, M., Müller, P., Klipp, W., Priefer, U., Simon, R. and Weber, G. (1984) In Veeger, C. and Newton, W.E. (eds), Advances in Nitrogen Fixation Research, pp 609-619, Martinus/Junk Pudoc, The Hague
12. Ausubel, F.M., Buikema, W.J., Earl, C.D., Klingensmith, J.A., Nixon, T. and Szeto, W.W. (1985) In Evans, H.J., Bottomley, P.J. and Newton, W.E. (eds), Nitrogen Fixation Research Progress, pp 165-171, Martinus Nijhoff Publishers, Dordrecht/ Boston/Lancaster
13. Scott, D.B., Court, C.B., Ronson, C.W., Scott, K.F., Watson, J.M., Schofield, P.R. and Shine, J. (1984) Arch. Microbiol. **139**:151-157
14. Fischer, H-M., Alvarez-Morales, A. and Hennecke, H. (1986) EMBO J. **5**:1165-1173
15. Ruvkun, G.B., Sundaresan, V. and Ausubel, F.M. (1982) Cell **29**:551-559
16. Corbin, D., Barran, L. and Ditta, G. (1983) Proc. Natl. Acad. Sci. USA **80**:3005-3009
17. Earl, C.D. and Ausubel, F.M. (1985) In Ausubel, F.M. and Szalay, A.A. (eds), Proc. 2nd Int. Sym. on Molecular Genetics of the Bacteria-Plant Interaction, Cornell University, Ithaca
18. Fuhrmann, M., Fischer, H-M. and Hennecke, H. (1985) Mol. Gen. Genet. **199**:315-322
19. Roberts, G.P. and Brill, W.J. (1981) Ann. Rev. Microbiol. **35**: 207-235
20. Dixon, R.A. (1984) J. Gen. Microbiol. **130**:2745-2755
21. Rossen, L., Ma, Q.S., Mudd, E.A., Johnston, A.W.B. and Downie, J.A. (1984) Nucleic Acids Research **12**:7123-7134
22. Lankhorst, R.K., Hontelez, J.G.J., van de Greef, W., van Kammen, A. and van den Bos, R.C. (1985) In Evans, H.J., Bottomley, P.J. and Newton, W.E. (eds) Nitrogen Fixation Research Progress, p 187, Martinus Nijhoff Publishers, Dordrecht/Boston/Lancaster
23. Schetgens, R.M.P., Hontelez, J.G.J., van den Bos, R.C. and van Kammen, A. (1985) Molec. Gen. Genet. **200**:368-374
24. Downie, J.A., Ma, Q.S., Knight, C.D., Hombrecher, G. and Johnston, A.W.B. (1983) EMBO J. **2**:947-952
25. Maniatis, R., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, a laboratory manual. Cold Spring Harbor Laboratory, New York
26. Beringer, J.E. (1974) J. Gen. Microbiol. **84**:188-198
27. Simon, R. (1984) Mol. Gen. Genet. **196**:413-420

28. Priefer, U. (1984) In Pühler, A. and Timmis, K.N. (eds) Advanced Molecular Genetics, pp 14-25, Springer-Verlag, Berlin/Heidelberg/New York/Tokyo
29. Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.W. and Ausubel, F.M. (1982) J. Bacteriol. **149**:114-122
30. Priefer, U., Simon, R. and Pühler, A. (1984) In Pühler, A. and Timmis, K.N. (eds) Advanced Molecular Genetics, pp 190-201, Springer-Verlag, Berlin/Heidelberg/New York/Tokyo
31. Priefer, U., Simon, R. and Pühler, A. (1985) J. Bacteriol. **163**:324-330
32. Buikema, W.J., Long, S.R., Brown, S.E., van den Bos, R.C., Earl, C. and Ausubel, F.M. (1983) J. Molec. Appl. Genet. **2**: 249-260
33. Southern, E.M. (1975) J. Mol. Biol. **98**:503-517
34. Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. **113**:237-251
35. Jagadish, M.N. and Szalay, A.A. (1984) Molec. Gen. Genet. **196**:290-300
36. Burkardt, H.J. and Pühler, A. (1984) In Methods in Microbiology, Vol. **17**, pp 133-155
37. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA **74**:560-564
38. Hong, G.F.J. (1982) J. Mol. Biol. **158**:539-549
39. Martinez, H.M. (1983) Nucleic Acids Research **11**:4629-4634
40. Sobel, E. and Martinez, H.M. (1986) Nucleic Acids Research **14**:363-374
41. Drummond, M., Whitty, P. and Wootton, J. (1986) EMBO J. **5**: 441-447
42. Hombrecher, G., Gotz, R., Dibb, N.J., Downie, J.A., Johnston, A.W.B. and Brewin, N.J. (1984) Molec. Gen. Genet. **194**:293-298
43. Sundaresan, V., Jones, J.D.G., Ow, D.W. and Ausubel, F.M. (1983) Nature **301**:728-732
44. Beynon, J.L., Cannon, M.C., Buchanan-Wollaston, W. and Cannon, F.C. (1983) Cell **34**:665-671
45. Better, M., Ditta, G. and Helinski, D.R. (1985) EMBO J. **4**: 2419-2424
46. Ow, D.W. and Ausubel, F.M. (1983) Nature **301**:307-313
47. Ow, D.W., Xiong, Y., Glu, Q. and Shen, S-C. (1985) J. Bacteriol. **161**:868-874
48. Buck, M. (1986) J. Bacteriol. **166**:545-551
49. Alvarez-Morales, A., Betancourt-Alvarez, M., Kaluza, K. and Hennecke, H. (1986) Nucleic Acids Research **14**:4207-4227
50. Buck, M., Miller, S., Khan, H., Drummond, M. and Dixon, R. (1986) Nature **320**:374-378
51. Kim, C-H., Ditta, G. and Helinski, D.R. (1985) In Evans, H. J., Bottomley, P.J. and Newton, W.E. (eds) Nitrogen Fixation Research Progress, p 186, Martinus Nijhoff Publishers, Dordrecht/Boston/Lancaster
52. Long, S.R., Buikema, W.J. and Ausubel, F.M. (1982) Nature **298**:485-488
53. Brewin, N.J., Wood, E.A., Johnston, A.W.B., Dibb, N.J. and Hombrecher, G. (1982) J. Gen. Microbiol. **128**:1817-1827
54. Simon, R., Priefer, U. and Pühler, A. (1983) Biotechnology **1**: 784-791
55. Rossen, L., Johnston, A.W.B. and Downie, J.A. (1984) Nucleic Acids Research **12**:9497-9508
56. Vieira, J. and Messing, J. (1982) Gene **19**:259-268