Organization and partial sequence of a DNA region of the *Rhizobium leguminosarum* symbiotic plasmid pRL6JI containing the genes *fixABC*, *nifA*, *nifB* and a novel open reading frame

Petra Grönger, Sundaram S.Manian¹, Helmut Reiländer, Michael O'Connell², 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 fixABC and nifA genes of the Rhizobium leguminosarum symbiotic plasmid pRL6JI have been identified. DNA sequencing of the region con-taining nifA 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 nifA gene. Comparison of three nifA proteins (Klebsiella pneumoniae, Rhizobium meliloti, Rhizobium leguminosarum) revealed only a weak relationship in their N-terminal regions, whereas the C-terminal parts exhibited strong Sequence analysis also showed that the R.leguminosarum homology. nifA gene is followed by nifB and preceded by fixC 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 *nifL* gene product (1,2) while the *nifA* protein is required for transcriptional activation of all *nif* operons, except its own, *nifLA* (3,4,5,6).

Within the last few years it has become obvious that the nifA 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 nifA 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 nifA have also been identified in Rhizobium trifolii (13) and Bradyrhizobium japonicum (14). The R.meliloti nifA gene, also designated fixD, is located between

Nucleic Acids Research

the fixABC operon and the nifB gene and transcribed in the same orientation (8,9,10,11,12,15,16,17). The function of the fixABC genes, which have also been localized in Bradyrhizobium japonicum (18), is still unknown. The nifB 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 nifA and nifB have also been identified in R. leguminosarum. In strain 248, a region of the symbiotic plasmid pRL1JI was shown to contain a K. pneumoniae nifB-like gene, called fixZ, as well as the last portion of an open reading frame homologous to the 3' end of K. pneumoniae nifA (21). For R. leguminosarum strain PRE, a nifA homologous gene has been identified on a 3.3 kb BamHI fragment and shown to be necessary for the expression of the nitrogenase structural genes nifHDK. Hybridization to the fixZ region was observed next to the nifA 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 nifA appears to be present on the R. leguminosarum symbiotic plasmids pRL1JI and pPRE and to map, analogous to R. meliloti, between nifB 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 fixABC*, *nifA* and *nifB* 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 nifA* 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	53
	B151	128C53 Cm, cured of	53
	B164	pRL6JI 128C53 Sm, contains pIJ1008	Brewin
R.meliloti 2011		wild type, Sm	Dénarié
Vectors			
pSUP202		pBR325 derived	54
pSUP205		derivative of pSUP202,	54
pSVB20		contains cos-site in Ap sequencing vector, based on pUC8 (56)	Arnold
Plasmids			
PIJ1008		recombinant between pRL6JI and pVW5JI, Nod, Fix, Hup, Tra, Nm (Tn5)	53
Hybridization Probes	l		
pRmSL26		Tc (pLAFR1), nodDABC and	52
pRmR3		Te (pACYC184), fixABC and	15
pRmW53 pRmW54		Te (pACYC184), nifK Te (pACYC177-Cm), nifA and most of nifB	Weber/Pühler 10
pMA152		Te (pACYC184), nodABC and	Aguilar/Pühler
pMA14.1		part of nodD Ap (pACYC177), nodC	Aguilar/Pühler

Media and growth conditions

ъ

4

B.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 *B.coli*: tetracycline 5μ g/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).

t

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. EcoRI 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 BCORI 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 *BcoRI* and all fragments were individually subcloned into the vector pSUP202. Restriction maps of all the *BcoRI* fragments were generated for the enzymes *BamHI*, *BglII* and *XhoI* 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.melilo-ti* 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 BamHI 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), *fixABC/nifA* (pRmR3) and *nifK* (pRmW53) probes suggested that two of the cosmids (Cos49, Cos56) contained the *nifHDK* region, while one cosmid clone (Cos4) carried the *nodDABC* and *fixABC/nifA* 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 EcoRI digest of Cos4 was probed with pRmR3 containing the R.meliloti fixABC/nifA region and with pRmW54 carrying the R.meliloti nifA/nifB genes. The hybridization patterns obtained, revealed that sequences homologous to the R. meliloti fixABC/nifA probe resided on a 5.45 kb EcoRI fragment, while homology to nifA/nifB was found to both a 5.45 kb and a 1.80 kb BcoRI fragment (data not shown). Similarly, both BcoRI fragments hybridized to the 1.5 kb BamHI fragment of pRmW54, which contains only nifA specific sequences. Hybridization experiments with plasmid pMA152 established that sequences homologous to the R. meliloti nodDABC region resided on two EcoRI fragments of sizes 2.95 kb and 1.35 kb; homology to nodC (probe pMA14.1) was only present on the 2.95 kb EcoRI 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 fixABC/nifA and the nifHDK regions.

Comparative analysis of the R. leguminosarum and R. meliloti fixABC and nifA regions by heteroduplex studies

To compare the structural organization of the fixABC/nifA containing regions in R.leguminosarum and R.meliloti heteroduplex molecules between the 5.45 kb BcoRI 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=BamHI, B'=BglII, C=ClaI, E=EcoRI, H= HindIII, P=PstI, S=SalI, S'=SmaI, X=XhoI.

kb EcoRI 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 fixC, the long segment could be readily assigned to the coding region for fixABC. However, the homology clearly extended beyond this region. The second segment of homology appeared within the nifA coding region but did not cover the entire fraction.

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

DNA sequence analysis of the R.leguminosarum nifA region

To better understand the dissimilarities within the nifA region of *R.leguminosarum* and that of *R.meliloti*, the 3.3 kb BamHI fragment of Cos4, known to start within the region homologous to fixC and expected to contain the entire nifA gene of *R.legumino*sarum (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 fixABC/nifA regions of *R.leguminosarum* and *R.meliloti*. The electron micrograph shows the actual heteroduplex structure obtained between the 4.8 kb *BcoRI* fragment of pRmR3 (15) containing the *R.meliloti* fixABC and part of the *nifA* gene and the 5.45 kb *BcoRI* fragment of Cos4 (compare also Fig. 1); the *R.meliloti* fragment was marked with transposon Tn5, inserted into fixC. The schematic representation shows the interpretation of the hybrid molecule, aligned to the maps of the respective *BcoRI* fragments. Restriction sites are as in Fig. 1.

tained for the 3.3 kb BamHI 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 gcoRI 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=BcoRI, S=SaII, 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 nifA 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) nifA 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 nifA proteins for the various domains are summarized in Table 2 which also includes the comparison to the K.pneumoniae ntrC sequence (9,41). The first 140 amino acids of the R.leguminosarum nifA 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 *nifA* proteins and is also conserved in ntrC. 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

G S E I K E Y A A E L I P E G C P
COA THE GAG ATE ANA GAG THE GEE GEA CAT CTT ATE COT GAG GEA CGE TTE 51
ALE GEA ATE CEA CAE ETE TIT SEE AAC GEE TEE ETE ETE SEE CAE CEE POR 105
CAN TTA ANC ANC GOC GTG CAT AGG GAG GGA TCA ANC CTT GOG ATG ACA TCA GGC 159
L H A G E A I F Q I K S R G G L H T <u>CTC ATG GCG GGT GAA GCG ATC TTC CAG ATA AAG AGC CGT GGC GGT CTC ATG ACG</u> 213
R R N L S L Y K G H L G R S F V H R ANG CGC NAT CTC TCT TAT ANG GGC ATG CTG GGT ANG TCG TTC GTC ATG ANA 267
D L H K H K D L P S L L H T D S H H GAC TTG ATG AAA CAC AAA GAT CTT CCA AGC CTC CAC ACC GAC AGT CAC AAT 323
F F H T Y P T L I S Q A A Q H F V B TIT TIC ATG AGG TAT CCA AGG CTT ATA TCT CAG GCG GCT CAA AAT TTT GTG CGC 375
V D G A P K I B T E K A T A A S F I GTC GAC GGT GGA CGT AMA ATC AMC AGG GAG AMG GCC ACA GCT GCC TCC TTT ATC 429
R A R S R W G L I S D A V R S A V S ANC GCA CGA TCC CGT TGG GGG TTG ATT AGC GAC GCG GTC CCC TCC GCC GTA TCT 483
W R OC TGG COT TAA AAGGAAAATTCO ATG AAG GCG ACC ACC ATT GAG CGC ATT GAG GAT 537
R L Y Q R R Y L V D T R R P H I T V ANG CTG TAC CAA AAC OGA TAT CTC GTC GAT ACT AGA OGC CCA CAC ATT ACC GTG 591
R P H R S P S P S L L A L T Q I C P COG CCG CAT CGG TCG CCA AGC CTA AGC CTG CTG ACG CAA ATC TGT CCG 645
A K C Y K V N K I G Q V A I V S .D G GCC ANA TCC TAC GAG GTG AAC GAA ATT GGT CAA GTG GCG ATT GTT TCG GAT GGC 699
C L E C G T C R V L A E A S G D I K TCC TTG GAA TGC GGC ACA TGC AGA GTG TTA GCC GAA GCG ACT GGC GAC ATA AAG 753
W H Y P R G G F G V L F K F G OP TIGG AAT TAT CCC CGG GGA GGG TTC GGG GTC CTC TTC AAA TTC GGA TCA GGAGTCC 808
CTACCTCCSGCSGGATAGCAGCGATCGATCCAGTCGCATTAGCSGCGGCGATTGAATGTGCATCCACTCAA 879
CCTITICEAAAGGCTAATTTICEAATTACAGGTATTTCTAACCTAGGAATATGCGTTAGCATATTTCGCGCAC 950
CAGGETAAAAAGCAGTGCACCCCCCCCTGTTTGTAGGC ATG ATT AAA CCA GAG GCG CGG CTC 1012
BILYDISKELISSFPLDM CAT ATT CTC TAC GAC ATC TCC AAA GAG CTT ATC TCT TCT TTT CCT CTA GAC AAC 1066
L L K A A N N A L V E H L R L R D G TTG CTG AAG GCT GCC ATG AAC GCC CTC GTC GAG CAT CTG CGA TTG CGC GAC GGC 1120
G I V I E G S G G E P W I R V R A P GGA ATC GTG ATT CAC GGC TCC GGA GGA GGA GAG CCT TGG ATA AAC GTA CGG GCT CCC 1174
I G D D V R S R S L T I B Q A D A I ATT GGG GAC GAC GTT CGC TCA CGT TCT CTG ACG ATT GAA CAG GCG GAC GCA ATA 1226
H R V I A S G S K H P G K H S V V L ANT COT OTC ATC GOT AGG GOT GAG AAG CAC TTT GOG AAA AAT TET GTC GTT CTC 1282
P V E V H R E A I G A L W I D P A Q CCC GTT ANA GTA ANC CCG ANA GCA ATC GGC GCG TTQ TGG ATT GAT TTC GCG CAG 1336
K & G A Q D E S L L A N I A V L I G AAA AGE GGA GET CAG GAA GE CTT ETG GEA ATG ATT GET GTE CTG ATE GGE 1390
L T C Q R D R B L C B D G G B V A B TTA ACC TOC CAG COC GAT COC GAA TTG TOC ACC GAC GGC GGC TCA GTC GCC GAG 1444
E Q Q A G Q I P R I R P R P B P T Q GAA CAA CAA GCG GGA CAG ATT CCG AAA ATC AAG CCG AAG CCT CAC CCC CAA 1498
L D K I D W I V G E S P A L K R V L CTC GAT AAA ATC GAC TOG ATC GTC GGG GAG AGC CCC GCG CTC AAG AGG GTA TTA 1552
A T T K I V A A T B S A V L L R G E GCC ACC ACC ACC CC GCG GTG CTC TTG AGA GAG GAG (1606
B G T G K B C F A R A I H A L S I R AGC GGC ACT GGC ANG GAG TGC TTT GCA AGA GCA ATA CAC GCG TTA TGG ATA CGG 1660
K S K A Y I K L S C A A L S S T V L MAA AGC AMG GOG TIT ATT AMG TTO AAT TOC GCC GCG CTG TOG GAA ACC GTT CTG 1714
E S E L Y G H E K G A Y T G A L L Q GAA TCC GAA TTG TTT GEC CAT GAQ ANG GEC CET TTC ACT GGC GET CTC CAT 1768
R A G R F S L A N G G T L L L D S : CGA GCT GGA CGT TTC GAG CTG GCC AAT GGC GGA ACG CTA TTG CTT GAT GAA ATT 1822
R A G R F E L A H G G T L L L D E 1 COA GOT COA COT TTC CAG CTG GCC AAT GCC GGA AGG CTA TTG CTT GAT GAA ATT G D V S P Q F Q A K L L R V L Q E G G C CAT GTA TCA CAC GCC AAG TCA TTG CCC GTA TTG CGC GAA GCC 1875
R A G R F S L A H G G T L L L D S : CGA GCT GGA CCT TTC CAG CTG GCC AAT GGC GGA ACG CTA TTG CTT GAT GAA ATT 1822 G D V S F Q F Q A K L L R V L Q E G GGC GAT GTA TCA CCA CAA TTC CAG GCG AAG TTA TTG CCC GTG TTA CAG GAA GCC B F B E L G G T K Y L K V D V R V X G G T G T C A G A G A G A G C T G G G G T A T T C A G G G G T T T C A G G G T G T G T T T T G G G G T T T T T G G G T T T T T T G G G T T T T T T G G T T T T T T T T T T
$\begin{array}{cccccccccccccccccccccccccccccccccccc$

D L Y Y R I H V V P I I L P P L R Q GAC CTC TAT TAC CGG ATC AAT GTG GTG CGG ATC ATT TTG CGG CCA CTT CGG CAG	2038
B D G D I S L L A Q V F L B Q F N N COC GAC GGA GAC ATT TCG CTT CTA GOG CAA GTG TTC CTC GAG CAA TTC AAC AAC	2092
A N D A N C D P G P S A I D I L S K GCA AAT GAT CGA AAT TGC GAC TTC GGG CCG TCG GCA ATA GAC ATT TTG TCG AAA	2146
C A F F G N V R E L D H C V Q R T A THE GCC TTC CCC GGC AAT GTT CGC GAG CTG GAC AAC THE GTT CAA AGG ACC GCC	2200
T L A S S N T I T S S D F A C Q Q D	
ACT CTC CCC ANT TCA ANT ACC ATC ACT TCA TCG CAT TITT CCC TCT CAG ENA CAC	2254
THE TOT TOT TOT TOT OUT OT	2308
A H H G L H S R D T H S G G L C A H GCG ATG AAT GGT CTC AAC TCG CGA GAT ACA ATG TCG GGC GGA CTG TCT GGC GAC	2362
A G T P E G A A A T I E A A G L T E GCA GGC ACT CCC AGC GGT GCC GCA GCC ACA ATC GAG GCA GCG GGC CTC ACC GAG	2416
R D R L I X A H I R A G W V Q A X A COT GAT CCG CTG ATC ANG GCA ATG CAG AGG CCT GGT TGG GTA CAG GCC ANA GCG	2470
A R I L G K T P R Q V G Y A L R R E GCT CGT ATC CTG GGT AMA ACG CCG CGG CAG GTC GGC TAT GCG CTA CGC CGG CAT	2524
R I D V K K E OP CCT ATC GAT GTG ANG ANG GAG TGA CAGCGATCGCCAAGAACTCCCCTAAGCGTGGCCCGAACGAC	2587
TEGGGEGGATCTATTTCATTTCTACAAGACATCTCCGGCAGCAAGCGGGAAACGAGCGGTAAGGCGAC	2658
ATANGCGCTGAAACAATAATATCTCATCGACCGGAATCTCTCTTGCTTTTGGAGCTGTAACCTTTCCTCAA	2729
CHGTGGAACHGTGGGATGTGGGAACCGGGAAATAAAGGTCGGGACGACCAGCAATGCCCCCCCC	2800
H P G R A S S Y G L S	
SCOCATESCCCCECT ATE CCC SET SEC CEC SCA TCT TCS TCC TAT SEC CTT TCS	2854
V T D D K D A R I W E R I K D H P C <u>GTG ACG GAT GAC AAA GAT GCG CGG ATC TGG GAG AGA ATT AAA GAT CAT CCC TGC</u>	2908
F E E Q A E E Y F A R H E V A V A F TIT TCA GAG CAA GCC CAT CAC TAT TTC GCT CGC ATG CAT GTC GCG GTC GCG CCT.	2962
A C M I Q C M Y C M R X Y D C T M R GCC TGC AAC ATC CAG TGC AAC TAC TGC AAT CGC AAA TAT GAC TGC ACC AAC GAA	3016
S R P G V A S V R L T P D Q A L R R AGC COT CCC GGG GTC GCA TCA GTA AAG CTA ACT CCC GAG CAG GCC CTA CGC AAG	3070
V L A V A S K V P E L S V I G V A G	3124
P G D A C Y D W R R T A A T P E G V	
CCG GOC GAC GCT TGT TAC GAC TGG AGG ANA ACA GCA GCG ACG TTT GAA GGA GTT	3178
A R E I P D I R L C I B T N G L A L acg aga gaa ata cet gac ate ana eta tec ate tec aca aat gga tig geg ett	3232
P D H V D E L A D N N V D H V T I T CCG GAC CAT GTC GAT GAG CTA GCT GAC ATG AAC GTC GAT CAC GTG ACG ATG ATG ATG	3286
T N N V D ATC AAC ATG GTG GAT	3301

Figure 4. Complete sequence of the 3.3 kb BamHI fragment containing nifA and adjacent regions. The 3.3 kb BamHI fragment, known to start in a fixC homologous region and expected to contain the entire nifA 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 *B*coRI 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 nifA* gene.

Localization of the R.leguminosarum fixC and nifB genes

Upstream of nifA, two open reading frames were localized,

	10	20 A	30	40	50	60
К.р.	MIHKSDSDTTVR	RFDLSQQFTANC	RESVERA	TEASKTIDEVL	SVIHNDAFMOH	GHICL
R.m.		MRKQDKRSAEIY	STERALMAP	TRLETTINNEV	NTUSLILRARR	GCLEI
R.L.		MIKPEARLHIL	DISKELISS	PPLONIIKAAM	NALVEHLRLRD	GCIVI
	70	80	 %	100	110	120
К.р.	YDBOORTISTEA	LOGTEDQTLPGS	TQI-RYRPG	EGLVGTVLAO	QSLVLPRVADD	ORPLO
R.m.	PASEGETKI	TAATRNSGS	PSAADYTVP	KAAIDOVHTAG	R-LVVPDVCNS	ELFKD
R.L.	HGEGGEPWI	+ NVRAPIGDD\	* /RSRSLTIEQ	ADAINRVIASC	E-KHFGKNSVV	LPVKV
		140	150	UU 160	170	190
К.р.	RLSLYDYDL-P	INVPLNGPHSRI	TCVIARHAN	ARQEERLPACT	RFLETVANLIA	OTIRL
R.m.	QIKWRGIGPTAF	AAAVEVDHET-	GGHLWPECA	EESDYDYEEEV	HFLSMAANLAG	RAIRL
R.L.	NRKAIGALWIDE	AQKSGAQDE		VLIGLTCORDR	ELCSDGGSV	
	190	C200	210	₂₂₀ D	230	240
К.р.	MILFTS	AADAPODSPRIM	RPRACTPSR	GFGLENMVGKS	PAMRQIMDIIF	QVSRW
R.m.	HRTISRRERTFA	EEDOEODNSRDI	OSOSSAROF	LLKNDGIIGDS	TALMTAVOTAP	WHAT
R.I.	<u></u>	EECOAGOIPKI-	КРКРНРТО-	LDKIDWIVCES	PALKRVLATT	IVAAT
K -	250	260	270			300
к.р.	DTTVLVRGESGT	GKELIANAIHHI *** = ==	SPRAAAAFV * * i*	KENCAALPDNI ***	LESELFGHEKG	APTGA
R.m.		GKECFARITHOI	STROKKPHI	KENCPALSESI	LESELFGHERG	AFTGA
R .(.	NSAVILINGESCI	GKECFARATHAI	Sterring	KUNGALISETV	Leselfgheko	AFTGA
Ka	310 10000000000000000000000000000000000	320	330	340	350	360
п.р.			*****	* 19*** * **		
R.m.	IADRVGRFESAN	GGTLULDEIGE *****	IPPAPQAKLI	RVIDEGEFERV	GGITKITLKMDVE ****	
R.I.	LIDRACRFEIAN	CGTUULDEIGD	/SPOPQAKLI	RVLDEGEFER	COLKLINNDA	Meri
Kn	370 जिन्द्राजी जिन्द्र सिंह की जिन्द्र	380 380	<u>390</u> 1977 - 1977 - 1977		410 10 10 10 10 10 10 10 10 10 10 10 10 10 1	420 1971 197
п.р.			* ****	HFILL		
R.M.			VECULPPLK:		** ** *	
R.L.	NKNLEVAVLEG	FRADLYYRINV	ALL IL PARK	BDDDISTITYD	FLEQ-FNNANI	FINCD
K n.	430 SDCATRTNMEVS	440	450 LERSAVUSE	460 CTADRD	470	480
D m				H H	TOOCPESDIM	CVHCS
п.ш.		******		*****		
R .I.,	GERMIDIUSKCA	PERVICELENC	Contraction of the second s	MINITER PROPERTY AND	QUQUSSALLING	MDGDG
	490	E 500	510	520	530	540
К.р.		VILPNHRDN	PFXALASSC: *** * *	*		KAGWV
R. m.	HGHIEIDAPAGT	TFLLGAPANDV	PREPCSAG	VASNI	IERORLISALE	ENGWN
R .(.	IGNDAMNGLNSR	THEGELCAHAG	TESGAAATI	EAACT	TERDRLIKAME	RAGNY
K	550	560	-			
к.р.	*****		HPRL *			
R.m.	QAKAARIILEKTP	ROVGYALRRHGV	UVRKL			
R.I.	OAKAARIKGKIP	ROVCHALRRHRI	đ vkke			

Figure 5. Comparison of the expected nifA amino acid sequence of R.leguminosarum (R.1.) with the nifA 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 nifA* amino acid sequence to nifA of *R.meliloti* (9,10) and nifA and ntrC 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 ntrC (41).

Domain	Coordinates	Number of residues compared	Percentage R.meliloti nifA	of homology to K.pneumoniae nifA	K.pneumoniae ntrC
A	1-184	141	24.1	11.4	n.d.
с	193-213	19	31.6	26.3	15.8
D	219-460	240	72.9	55.8	44.6
Е	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 fixC, since from minicell experiments, the fixC 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* fixC coding region and was compared with the 3' end of the *R.meliloti* fixC 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 nifA 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 nifB locus, since it had been shown that in *R.leguminosarum*, as well as in *K.pneumoniae* and *R.meliloti*, the nifA gene is neighboured by a sequence homologous to nifB (21,23). The amino acids deduced from this sequence were therefore compared to the published sequence of fixZ, the nifB-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 *nifB* sequence (pRL6JI) to the *fixZ* sequence of pRL1JI (21). Asterisks denote identical residues; arrows indicate presumptive reading frame shift errors in the *fixZ* 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 nifB* gene (H.R., manuscript in preparation), also in those sections differing from the *fixZ* sequence (124 amino acids being identical out of 162 residues compared = 76.5%). Putative frame shift errors in the *fixZ* sequence, as indicated in Figure 6, might be responsible for the divergency between the two *R.leguminosarum nifB* genes.

Identification of a novel open reading frame preceding the nifA 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 fixC and the nifA gene. The heteroduplex experiment between the fixABC/nifA regions of R.leguminosarum and R.meliloti (Fig. 2)

R.m. MKTAMPERIEDKLYQNRYLVDAGRPHITVRPHRSPSINLLALTRVCPAKCYELNETGQVE

R.M. VTADGCMECGTCRVLCEANGDVEWSYPRGGFGVLFKFG R.I. IVSDGCLECGTCRVLAEASGDIKWNYPRGGFGVLFKFG

Figure 7. Comparison of the amino acid sequences of the novel open reading frames located between the fixC and nifA 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 fixABC clearly extended beyond the fixC coding regions. This assumption was confirmed by sequence analysis, which revealed the presence of an open reading frame downstream of the fixC 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 fixC 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 fixABC, nifA and nifB 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 nifA homologue and the nodC 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 nifA homologous gene was found to be flanked on one side by fixZ, the homologue of the K.pneumoniae nifB locus, and by a cluster of fix genes on the other side (21, 23,24).

The R.leguminosarum nifA 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 nifA proteins in their C-terminal part which is also significantly homologous to ntrC, although to a lower degree. The two Rhizobium nifA 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),

Nucleic Acids Research

followed by domain E which contains a sequence related to a DNA binding motif (underlined region in Fig. 5). Like in K.pneumoniae nifA, 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 nifA proteins is less striking (domains A and C). The weak homology between R.meliloti and K.pneumoniae nifA in this area, already described by Drummond et al. (41), could be confirmed by including the comparison with R.leguminosarum nifA, the homology of which to K.pneumoniae is even lower. Although not necessarily expected, the two Rhizobium nifA proteins also exhibit only a weak relationship in their Nterminal 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 nifA (187 bp) and nifB (267 bp). Further, no homology to nifA or ntrC regulated promoter sequences (12,43,44,45,46,47,48) or to the R.meliloti nifA promoter (9) was detected. Also, no sequences resembling the upstream elements shown to be required for nifA activation in some species could be identified (6, 45, 49, 50). This is in contrast to R.meliloti where (i) nifA is preceded by a promoter sequence and can be expressed either from its own promoter or from the fixABC promoter p2 (9,51) and (ii) the intergenic region between nifA and nifB 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 nifB homologue fixZ could be found (21). These data suggest either, that in R.leguminosarum there are in fact no promoters in front of the nifA and nifB 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 nifA and parts of fixC and nifB, the sequenced fragment contained a novel open reading frame 5' to the nifA 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 *fixC* 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 *fixABC* 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 *fixABC*, *nifA* and *nifB* regions. Dissimilarities as they occur in the intergenic regions or in the N-terminal part of the *nifA* 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: ¹Department of Dairy and Food Microbiology, University College, Cork, Ireland; ²National Institute for Higher Education, Dublin, Ireland

*To whom reprint requests should be sent

REFERENCES

- Hill, S., Kennedy, C., Kavanagh, E., Goldberg, R. and Hanau, R. (1981) Nature 290:424-426
- 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
- Buchanan-Wollaston, V., Cannon, M.C., Beynon, J.L. and Cannon, F.C. (1981) Nature 294:776-778
- 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
- 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
- Downie, J.A., Ma, Q.S., Knight, C.D., Hombrecher, G. and Johnston, A.W.B. (1983) EMBO J. 2:947-952
 Maniatis, R., Fritsch, E.F. and Sambrook, J. (1982) Molecular
- 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