SYMBIOSIS (2009) 48, 110–119 ISSN 0334-5114

Sequence analysis of a DNA fragment from *Sinorhizobium fredii* USDA257 which extends the nitrogen fixation host range of *Rhizobium* species NGR234 to soybean, *Glycine max* (L.) Merr cultivar Peking

B. Boboye[1,2*], G. Nyakatura[4], A. Rosenthal[4], X. Perret[3], W.J. Broughton[3], and Th. Boller[1]

[1]Botanisches Institut der Universität Basel, Hebelstrasse 1, 4056 Basel, Switzerland;
[2]Current Address: Department of Microbiology, The Federal University of Technology, P.M.B. 704, Akure, Ondo State, Nigeria, West Africa, Email. boboye_b@yahoo.com;
[3]Laboratoire de Biologie Moléculaire des Plantes Supérieures, Université de Genève, 1 Ch. de l'Impératrice, 1292 Chambésy/Genève, Switzerland;
[4]Institut für Molekulare Biotechnologie, Abteilung Genomanalyse, 07745 Jena, Germany

(Received June 1, 2008; Accepted August 19, 2008)

Abstract
A fragment of DNA (pBTBX) from the genome of *Sinorhizobium fredii* USDA257 was sequenced by shotgun strategy to identify the potential genes which enabled the *Rhizobium* species NGR234 to fix nitrogen on soybean, *Glycine max* (L.) Merr cv. Peking. The total length of the cosmid is 32,824 base pairs with a GC content of 61%. A 29 open reading frames (ORF) were identified representing 71.8% (23,574 bp) of the cosmid. Out of these ORF, 96.5% (22,749 bp) were identical and similar to reported and hypothetical genes and proteins. The remaining 3.5% (825 bp) had no apparent similarity to any genes in the data base. Gene and gene products found on the DNA fragment include those involved in the synthesis of Fe-Mo component of nitrogenase, regulation of nitrogen fixation, transport of amino acids and sugars, chemotaxis and transcriptional regulation.

Keywords: Sequence analysis, DNA fragment, *Sinorhizobium fredii*, *Rhizobium* species NGR234, soybean Peking

1. Introduction

Generally, legumes fix nitrogen in symbiotic association with compatible bacteria collectively known as rhizobia. Rhizobia are grouped majorly into *Rhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Mesorhizobium* and *Sinorhizobium*. During the symbiosis, signals to and from both the macro- and micro-symbionts are released (Fischer, 1994). This exchange of chemical signals between soil bacteria (rhizobia) and legumes termed a molecular dialogue involves two main groups of molecules: *nod* gene-inducing flavonoids from plants and the mitogenic lipochito-oligosaccharide Nod factors (NFs) of rhizobia. The NFs synthesized by rhizobia elicit, at very low concentrations and in a specific manner, various symbiotic responses on the roots of the legume hosts (Debellé et al., 2001; Shaw et al., 2006; Steinkellner et al., 2007). This is because rhizobia respond to chemoattractants and growth-enhancing compounds in root exudates, and several plant non-flavonoids possess *nod* gene-inducing properties (Cooper, 2007). A number of nodulation genes which specify the synthesis of NFs have been identified. All rhizobia, in spite of their diversity, possess conserved *nod*ABC genes responsible for the synthesis of the N-acylated oligosaccharide core of NFs, which suggests that these genes are of a monophyletic origin. Other genes, the host specific *nod* genes, specify the substitutions of NFs (Debellé et al., 2001). Entry into the plant is restricted to bacteria that have the "keys" to a succession of legume "doors". Some symbionts intimately associate with many different partners (and are thus promiscuous), while others are more selective and have a narrow host range. This is related to difference in type of NF produced by rhizobia

*The author to whom correspondence should be sent.

Presented at the 15th International Congress on Nitrogen Fixation, January 21–26, 2007, Cape Town, South Africa

Table 1. Predicted genes and proteins encoded by pBTBX of *S. fredii* USDA257 which confers on NGR234 the ability to fix nitrogen on *G. max* cv. Peking.

Orf	Base posit. in plasmid	Funct. name	Frame	RBS' and start codon	Homolog gene or amino acid position in plasmid	Homolog name	Gene base/ amino acid length	Organism	Protein access no.	Identity (%)	Similarity (%)	Description
bt1	1-213		-3	gGGAGc-T-4ATG	34-186	BJORF	1447	BJ	M17635	67	67	DNA region with 3 ORF complete cds
					72-165	YBIO	786	*E. coli*	P75783	37	59	Hypothetical 86.8 kD protein in Ding-GLNQ intergenic region
bt2	294-539	*cat*	-2	cGGcGGc-9-ATG	330-509	SCRSCATC	212	*S. cerevisiae*	M12865	64	64	CAT repetitive element, clone pYCAT8
					369-522	NPT2	68	*Mus musculus*	Q62111	34	59	Na+.PO4 cotransporter type III fragment
bt3	1296-1691	*gpm*B	-3	cGccGGc-5-GTG	1357-1614	GPMB	349	*Syn.* sp.	P72649	35	54	Phosphoglycerate mutase
bt4	1631-4093		-1	cGGAGGT-4-ATG	1691-4079	?	821	*Syn.* sp	P74690	61	74	Hypothetical 92.4 kD protein
bt5	4491-4615		1	gcGAGaa-5-TTG	4495-4614	U650AH	55	*Myc. leprae*	Q50135	35	47	Hypothetical 5.8 kD protein
bt6	4672-5394		-3	AGGAccT-4-ATG	4708-5368	?	248	*Sal. typhi*	P74886	24	44	High affinity periplasmic glutamine binding protein
bt7	5397-7775	*nif*L	-2	gtGtGGT-10-ATG	6393-7689	NifL	840	*Syn.* sp.	P72843	36	55	Nitrogen fixation positive activator protein, NifL
bt8	8046-8954	*gcv*A	-2	AGGaaa-7-ATG	8076-8946	GCVA	305	*E. coli*	P32064	37	60	Glycine cleavage system transcriptional activator
bt9	9108-9767		3	AGGcGGc-6-ATG	9164-9671	SGKSACPG	5364	*Str. griseus*	X77865	62	62	Ketosynthase, Acyl carrier protein, ketoreductase, cyclase and dehydrase
		*nod*G			9108-9753	NODG	245	*R. meliloti*	P06234	33	52	NODG
bt10	10162 -10896	*lar*	1	gtGAGGT-13-ATG	10219 -10696	RNU87960	2479	*Ratus norvegicus*	U87960	62	62	Leukcoyte common antigen receptor (LAR) gene trans-spliced alternative untranslated exon
bt11	10969 -11562		1	AGGAGGg-3-ATG	11006 -11445	MCU60315	190,289	Virus	U60315	61	61	Mollusum contagiosum virus subtype, complete genome cds
bt12	11660 -13156		2	gGGAGGa-4-ATG	11672 -13136	RBSA	493	*H. influenzae*	P44735	37	59	Ribose transport ATP-binding protein
bt13	13173 -14237	*mxa*	3	gGGAGaa-4-ATG	13218 -14216	AF017434	4815	*Mt. extorquens*	AF017434	58	58	Methanol oxidation genes *mxa*E, *mxa*H, *mxa*B and *pmi*-like genes, complete genome cds
					13290 -14187	Y4mJ	333	NGR234	P55569	30	54	Predicted ABC transporter ATP-binding protein Y4mJ
bt14	14285 -15385	*mcp*	2	gGGAGGa-7-ATG	14285 -15197	RCMCPAB	5186	*Rh. capsulatus*	L48927	59	59	Methyl-accepting chemoreceptors (*mcp*A and *mcp*B) genes complete genome cds
					14444 -15305	Y4mL	324	NGR234	P55568	24	38	Predicted ABC transporter periplasmic binding protein Y4mL precursor
bt15	15431 -16450	*gnt*R	-1	AGGAGGa-4-GTG	15489 -16445	GNTR	331	*E. coli*	P46860	31	52	GNTR utilization system GNT-1 transcriptional repressor

Table 1. Continued.

Orf	Base posit. in plasmid	Funct. name	Frame	RBS[1] and start codon	Homolog gene or amino acid position in plasmid	Homolog name	Gene base/ amino acid length	Organism	Protein access no.	Identity (%)	Similarity (%)	Description
bt16	16741 -17544	*nif*E	1	AGGAGaa-3-ATG	16754 -17545	AVINIFE	1509	*Az. vinelandii*	X07293	56	56	*nif*E gene
					16774 -17383	Y4tE	300	NGR234	P55659	25	47	Predicted ABC transporter periplasmic binding protein Y4tE precursor
bt17	17612-18316		2	cGGAGtc-3-ATG	17672 -18284	Y4tF	238	NGR234	P55660	30	49	Predicted ABC transporter periplasmic binding permease Y4tF
					17663 -18284	Y4tG	231	NGR234	P55661	28	50	Predicted ABC transporter periplasmic binding permease Y4tG
bt18	18313-18972		1	cGGAGGT-6-ATG	18373 -18943	Y4tF	238	NGR234	P55660	30	51	Predicted ABC transporter permease protein Y4tF
					18331 -18952	Y4tG	231	NGR234	P55661	28	49	Predicted ABC transporter permease protein Y4tG
bt19	18953-19684		2	AGGAGGc-3-ATG	18965 -19673	Y4tH	257	NGR234	P55662	48	65	Predicted ABC transporter ATP-binding protein Y4tH
bt20	20003-20914		2	tatcGGT-5-ATG	20078 -20417	MJ0604	100	*M. jannaschi*	Q58021	34	50	Hypothetical protein MJ0604
bt21	21487-21918		1	cGGAGat-5-ATG	N.o.	N.o.	N.o.	N.o.	N.o.	N.o.	N.o.	Hypothetical 16.2 kD
bt22	22162-22341		-3	AGGctGa-6-TTG	22186 -22338	YDJA	183	*E. coli*	P24250	40	53	Hypothetical 20.1 kD protein in SELD-SPPA intergenic region (ORF 183)
bt23	22424 -23599	*sna*C	-1	cGGAGac-7-ATG	22746 -23509	SP21216	770	SPS	U21216	60	60	NADH-FMN oxidoreductase *(snaC)* gene, complete cds
					22835 -23597	?	351	Ps KHP41	005599	34	45	Integrase-like protein
					22772 -23594	RECR	343	BPP1	P06956	26	46	Recombinase CRE
bt24	23670-23948		3	AGGgGGa-8-ATG	23709 -23937	REPC	314	*Staph. aureus*	P03064	24	49	Replication initiation protein (protein A)
bt25	23991-24377		-2	AcGgGcT-9-ATG	23991 -24369	MTCY20 H10.05	131	Myc. tuber	P96914	61	75	Hypothetical 14.4 kD protein
bt26	24925-25317		-3	tGGcGGg-10-GTG	N.o.	N.o.	N.o.	N.o.	N.o.	N.o.	N.o.	Hypothetical 14.7 kD protein
bt27	25428-25991		-2	tGGActg-8-GTG	25428 -25977	Y4jO	321	NGR234	P55515	71	82	Hypothetical 36.1 kD protein Y4jO
bt28	25988-27091		-1	AtGAttc-9-ATG	26687 -27086	Y4jP	321	NGR234	P55515	61	75	Hypothetical 36.1 kD protein Y4jP
					25994 -26588	Y4jP	262	NGR234	P55516	63	73	Hypothetical 29.5 kD protein Y4jP
bt29	31132 -32253	*mcp*	3	AGccGGT-7-TTG	31465 -31699	MCP	99	Myc. tuber	E1191430	32	42	MCP protein

Posit., position; funct., functional; RBS, ribosome binding site; the number between RBS and start codon (for example 4 in gGGAGcT-4-ATG) indicates the position of RBS relative to the translation initiation start site. Az, *Azotobacter*; Bj, *Bradyrhizobium japonicum*; BPPI, Bacteriophage P1; E, *Escherichia*; H, *Haemophilus*; M, *Methanococcus*; Mt, *Methylobacterium*; Myc. tuber, *Mycobacterium tuberculosis*; NGR234, *Rhizobium* species NGR234; Ps, *Pseudomonas*; R, *Rhizobium*; Rh, *Rhodobacter*; S, *Saccharomyces*; Sal typhi, *Salmonella typhimurium*; SPS, *Streptomyces pristinaespiralis*; Staph, *Staphylococcus*; Str, *Streptomyces*; Syn, *Synechocystis*; N.o., None obvious.

(Perret et al., 2000; Debellé et al., 2001; Shaw et al., 2006; Steinkellner et al., 2007). At the onset of the association, exudates such as flavonoids (the strongest *nod* genes inducers) are secreted by plant roots into the rhizosphere.

Following root hairs colonization by the bacteria, common *nod* genes (*nod*ABC) are expressed in the microbial cells under the control of the regulatory *nod* genes (*nod*D, *syr*M or *nod*UVM) (Fellay et al., 1995; Hanin et al., 1998). This results in the biosynthesis of the lipo-oligosaccharide backbone of Nod factors which are

decorated by the products of host-specific *nod* genes; the action of which causes root hairs to deform, branch, curl, form infection threads.

Rhizobium species NGR234 and *Sinorhizobium fredii* USDA257 are both closely allied broad-host-range rhizobial strains. The former bacterium distinctly nodulates more than 110 genera of legumes as well as the non-legume *Parasponia andersonii* but induces non-nitrogen-fixing nodules on soybeans (Trinick, 1980; Lewin et al., 1987; Relic et al., 1994; Pueppke and Broughton, 1999). In contrast, USDA257 forms effective nodules on many cultivars of soybeans including cv. Peking (Keyser et al., 1982; Balatti and Pueppke, 1992).

Recently, we screened a genomic DNA library of USDA257 and identified a clone containing a DNA fragment, pBTBX which endows NGR234 with the ability to fix nitrogen on *Glycine max* (L.) Merr cv. Peking. Nodule-like structures formed by the wild-type NGR234 lack rhizobial cells in comparison with bacteroids-containing nodules produced by the transconjugant, NGR234(pBTBX) on the soybean. In order to identify and study the molecular basis of the genetic locus/loci on the pBTBX which confers this capability on NGR234, we have shotgun-sequenced the DNA fragment and complemented NGR234 with a gene segment from the pBTBX. Here, sequence analysis of the DNA fragment is presented.

2. Materials and Methods

Construction of shotgun library

Procedures employed in this work were according to Hanahan (1983) and Sambrook et al. (1989). Cosmid DNA (pBTBX) was prepared from a late log-phase grown cells of the clone containing the DNA fragment (obtained from *Sinorhizobium fredii* USDA257 which enabled NGR234 to fix Nitrogen). The DNA (pBTBX) was sonicated and ends of fragments were filled-in with nucleotides in the presence of T4 DNA polymerase. DNA fragments ranging in size from 1.6 to 3.0 kb were ligated to *Sma*1-cleaved M13mp18 at 16°C. Aliquots of ligated DNA were used to transform competent cells of DH5α. Recombinant phages were prepared from the transformed DH5α and used for sequencing.

Sequencing and nucleotides assembly

This sequence analysis was carried out at the Institut für Molekulare Biotechnologie, Abteilung Genomanalyse, Jena in Germany. Sequencing was carried out using Dye Terminator/Thermo Sequenase sequencing methods essentially as reported by Freiberg et al. (1996).

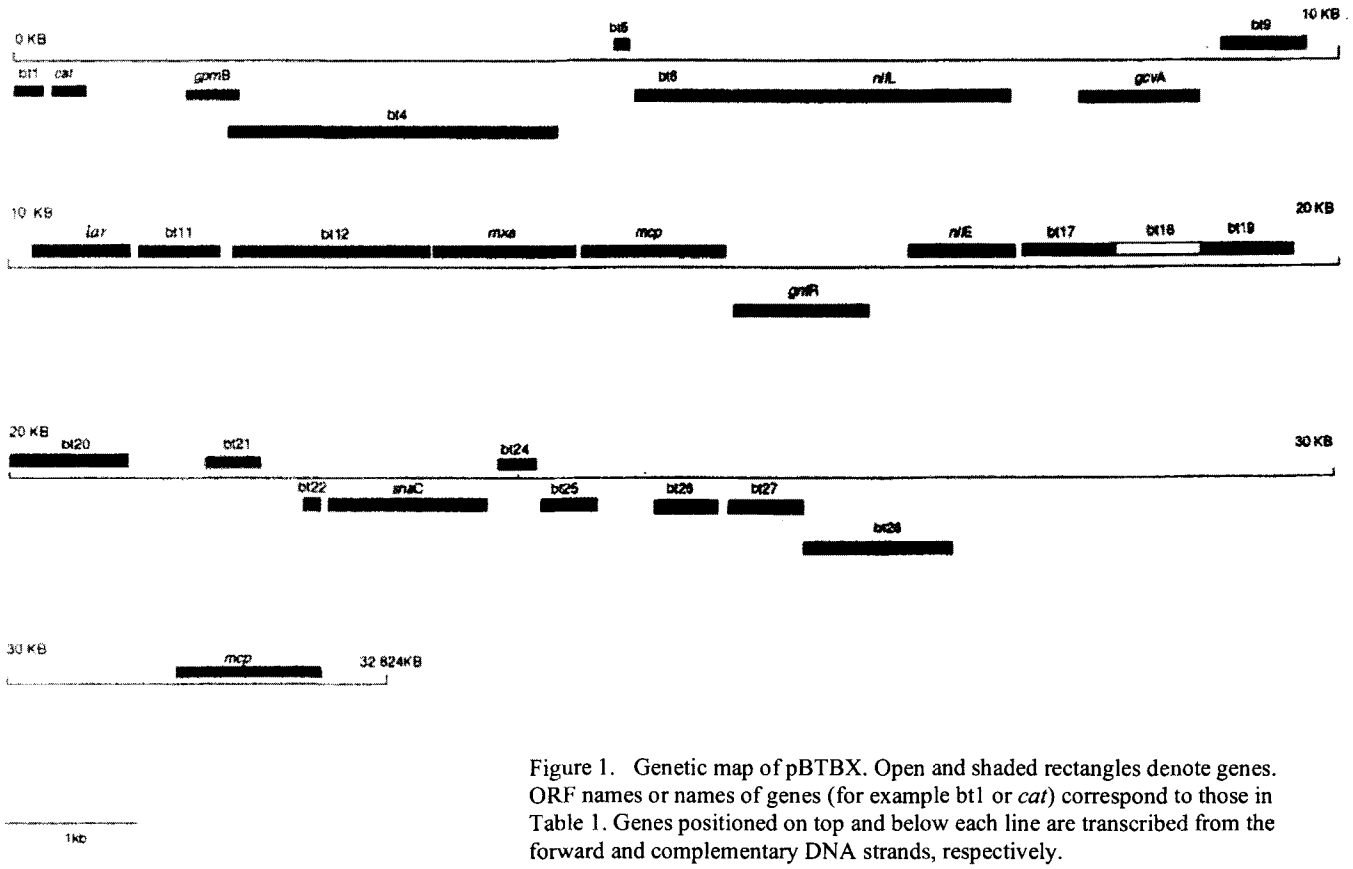

Figure 1. Genetic map of pBTBX. Open and shaded rectangles denote genes. ORF names or names of genes (for example bt1 or *cat*) correspond to those in Table 1. Genes positioned on top and below each line are transcribed from the forward and complementary DNA strands, respectively.

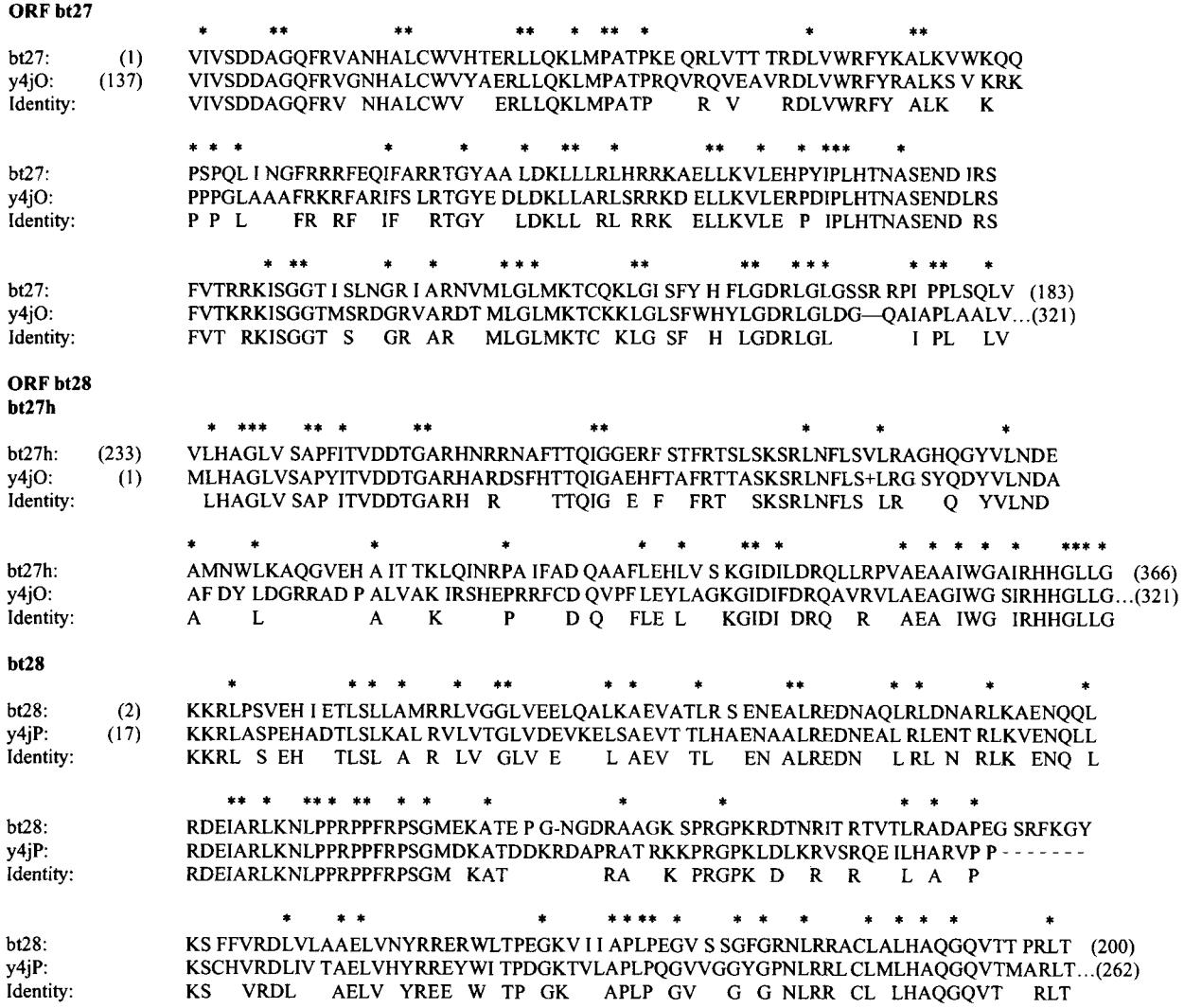

Figure 2. Alignments of protein sequences encoded by open reading frames bt27 and bt28 with y4j0 and y4jP of *R.* species NGR234. bt27h is bt27 homolog which exist in the same ORF with ORF bt28. Amino acid numbers in the protein corresponding to the first and last residues in the alignments shown are presented in parentheses. Amino acids identical in both proteins are shown below the alignments. Conserved leucine, alanine, glycine, proline and isoleucine are indicated with asterisks.

Nucleotides were compiled and edited after those for M13mp18 were removed. Gaps between contigs were closed with additional sequences obtained from dye primer and big-dye terminator sequencing methods.

Analysis of nucleotides

Open reading frames and intergenic regions were looked for in the nucleotides by using *Rhizobium meliloti* matrix of GeneMark prediction program version 2.3 (Borofsky, 1995). Search for similar and identical protein and gene sequences was conducted with the GCS computation program using the BLAST Network Service.

3. Results and Discussion

The DNA fragment (pBTBX) is 32,824 base pairs long with a GC content of 61%. Twenty nine Open Reading Frames were detected in the pBTBX (Table 1, Fig. 1). The initiation codon is ATG for most of the genes except for seven cases where GTG and TTG are used. The putative ribosome binding sites are generally rich in purines and showed variable locations in the sequence. Collectively, 72% (23,574 bp) of the pBTBX are potentially coding (Fig. 1). Two and half percent (825 bp) of the sequence show no significant similarity to any known genes or gene products in the database bank. Nucleotide content of the non-coding regions is essentially the same as that of the coding parts.

```
                                  -35
                                                        ***
btnifE:                            ATGACGATCTACCAAGTGGaggagaaACCATGACGATTTCG
                          -10

btnifE:   (13)AACTTTGTGCGGACAATGGT-GGC-A-GCCATCGGAATAGCG-GCG--GCCGGCCTCGTT
AzvnifE:(451)ACCTGCGTGCCGGCGCTGATCGGCGACGACGTCGACGCAGTGTGCAAAGCCGCCGCCGA-

btnifE:       GCGGCTGCGGAGGCGCAGGCCGCCACGCTGAACGACATCATCTCGCGCGGTACCGTGCGC
AzvnifE:      GCG-CTTCGGCACCCCGGTCATCC-CGGTCGACTCGGCCGGCTTCTACGGCACCAAGAAC

btnifE:       ATCGGAGTCCTGACCGGCGCGCCGCCCATGGGAATGGTCGACGAACAGGGCAACC-CAA-
AzvnifE:      CTCGGCAACCGCATCGCCGGTGAGGCCATGCTCAAGTACGT-GATC--GGCACCCGCGAG

btnifE:       CCGGCTACGACGTCGACGTGGCCAATCTGATTG-CCGGCTATCTGTC-GCT-GCCGGTCG
AzvnifE:      CCCGATCCGCTGCC--CGTCGGCAG-C-GAGCGTCCGGGCATCCGCGTGCACGACG-TCA

btnifE:       AGCTCGT-GCCGCTGACGCCGCCGGCT-CGCATTCCCGCTT-TGCAGAC-CGGCAAGGTC
AzvnifE:      ACCTGATCGGCGAGTACAACATCGCCGGCGAGTTCTGGCATGTCCTGCCGCTGCTCGA-C

btnifE:       GATTTCCTCGT-CGCCACC-CTGGCGCC-GACCGGGGA-GCGCGCCAAGACGGTGATGTT
AzvnifE:      GAACTGGGCCTGCGGGTGCTCTGCACCCTGGCCGGCGATGCGCGCTACCGCGAGG-TGCA

btnifE:       CACCCAGCCCTACAGCGCCTTCAACATGGACATCATCTCCG-GTCCCGAC-CA--GAAAT
AzvnifE:      GACCATGCACCGCGCCGAAGTGAACATGATGGTCTGCTCCAAGGCCATGCTCAATGTCGC

btnifE:       TTGCAAAGCTTGCCG-ATCTCGAAGGCAAGCGCGTCGC-CGTCAACCGTGGCT-CGTCG-
AzvnifE:      TCGCAA-GCT-GCAGGAAACCTACGGCACGCCC-TGGTTCGAGGGCAGCTTCTACGGCAT

btnifE:       CAG-GAGACGGCGCTGCGCAAGGCGGCAGTTC-CCGGCCTGGAAATCGTCG-TCTACGAG
AzvnifE:      CACCGACACCTCCCAG-GCGCTGCGCGACTTCGCCCGGCTGC---TCGATGATCC-CGAC

btnifE:       GATGATTCCACCAGCGCACAGGCGCTGATCGCCGGCCAGGTCGATGCGGTC-GCGCTGCC
AzvnifE:      C-TGA--CCGCCCGCACCGAGGCGCTGATCGCGCGCGAGGAGGCCAAGGTCCGCGCCGCC

btnifE:       CTCGA-CGGTCG-GTGAGGCGATCATCAAG-CAGCGCCCAGATGCCGGTCTGCAG-GTTG
AzvnifE:      CTCGAACCCTGGCGTGCG-CG-TCTGGAGGGCAA-GCGC-G-TGCTGCTCTACACCGGCG

btnifE:       GCTTCACCTTCTTCCAGCAGGGCAATTCGATGGCG-ACCCGGATGGAGGACTTTGAGATC
AzvnifE:      GCGTGAAGTCCTGGTCGGTGGTTTCCCCCCTGCAGGACCTGG--GCATGAAGGTG-G-TC

btnifE:       CGCCAGTGGCTCAACACCGCCATCTACCTGATGAAGATCTCCGGCG-ATCT-CGACAAGA
AzvnifE:      -GCCACCGGCACCA-A--GAAGTCCACC-GAGGAAGACAAGGCACGCATCCGCGAACTGA

btnifE:       TCG-CGACGAAGTGGACCGGC-CGCCCGATGCCGACGCTTCCGTCCTTCTGA (804)
AzvnifE:      TGGGCGACGACGTCAAGATGCTCGAC-GAGGGCAATGCGCGGGTGCTGCTGA (1243)
```

Figure 3. Nucleotide sequence of the *nif*E homologue (btl6*nif*E) and the homologous gene from *Azotobacter vinelandii* (Azv*nif*E). First and last nucleotides in the alignments shown are indicated in parenthesis. Positions of identical nucleotides are underlined. Putative ribosome binding site is shown in bold lower-case letters; hypothetical translation initiation codon is asterisked. Introduced gaps to give the best alignments are indicated with hyphens. Presumed Pribnow boxes are overlined.

The highest similarities found were with proteins of the closely related NGR234 (Table 1). The hypothetical proteins of ORFs bt27 and bt28 show strong identities of 71% and 63% and, similarities of 82% and 73%, respectively to the hypothetical y4jO and y4jP proteins of NGR234. Generally, the proteins are hydrophobic (Fig. 2). Thus, they may be important in the formation of cell structural components involve in transport systems. Other ORFs that have homology to transporting genes lie at 294 to 539bp (ORF bt2), 4672 to 5394bp (ORF bt6) and between bases 11660 and 19684bp (ORFs bt12, bt13, bt14, bt16, bt17, bt18 and bt19; Table 1). These are probably ATP binding cassette (ABC) type transporters except for bt2. The bt6, bt16, bt17, bt18 and bt19 show similarities ranging from 44 to 65% to the periplasmic glutamine binding protein of *Salmonella typhimurium*, y4tE, y4tF and y4tG, the ABC transporter binding protein and permeases of NGR234. The bt12, bt13 and bt14 show similarity matches to the ribose transport ATP-binding and ABC periplasmic binding proteins. The ABC superfamily is generally made up of multicomponent primary active transporters, capable of transporting both small and macromolecules in response to ATP hydrolysis (Paulsen et al., 1997). These transporters may influence nitrogen fixation considerably. Two genes (*prs*DE) coding for protein secretion which showed Fix^{-} defects in nodule bacteroids formed by *R. leguminosarum* bv. viciae and *R. leguminosarum* bv. trifolii in pea and vetch appeared to

```
bt7NIFL: (332): LAVMTDVT-ELKTAEQRNHVQAITDHLTGLLNRPGFEMALDAAIRQTAEDGGELACLLID
SynNIFL: (390): VAVKEDITKEKQQAEALFH-QAHYDHLTGLPNRILAKDRLQQAIESALRQKHIFGLMFLD

bt7NIFL:        LDRFKQINDNLGHAAGDEVLRQIAGRIRAQVRGEDKVGRLGGDEFVVLIPASKAQNAALQ
SynNIFL:        LDNFKKVNDTLGHDAGDQLLVEVSERLQRALRQTDTVARLGGDEFLIILDQVSHSRKLMA

bt7NIFL:        ISERIAAACAEPVIVDGNTLSLSASIGIALYPIQAITAAELLQKSDMAMYARKHNGKNGA
SynNIFL:        IAQRLLRVMRQPVNLQGLEFFVHGSIGITVFPDDGFHADVLLRNADTAMYAAKLAGRNMF

bt7NIFL:        KLFDPRMASLAQERLKIDTYIEEGLRQDWFDVHLQPIVDLKVGRIAGFEALMRLNHPEHG
SynNIFL:        RFFTPHMNQAAQQRMAIESELRQGLSRQEFQILYQPIVSLESGQIVGAEALMRWHNRLLG

bt7NIFL:        VLPPADIIRVAEETGAILRIGERIFEKAVAHLARLTSVPGLENAYLAVNFSPLQFCPK-L
SynNIFL:        TVPPDQFIPIAEEVGLIVELGEWLLDNVCCQAAHWHSALGEQTFWVSVNVSPRQLKDSYF

bt7NIFL:        PASTVSTLMKWGITPSRIVIEITEAVLMHHSPDIRDVLGALSSAGMKIALDDFGTGYSSL
SynNIFL:        VAILQGFLQRYQVRPEWLELEITENLILEENGDLLKNLSDLEEENIALSLDDFGTGYSSL

bt7NIFL:        SYLVHFPVNIIKIDQAFTRSLTDESEMVRRRVRKLVAGIHTVAKELNCQVVAEGIETEEQ
SynNIFL:        NYLRKFNFNSLKIDRSFVELLPHDNNTVGL-VRAIIAMAH----HLELKVIAEGIETPEQ

bt7NIFL:        LNALLSLDVNSGQGY (764)
SynNIFL:        WNFLRLQGCDYGQGY (818)
```

Figure 4. Comparison of amino acid sequence of protein bt7NIFL with the homologous protein (SynNIFL) from the nitrogen-fixing cyanobacterium, *Synechocystis* species. Amino acid numbers in the protein corresponding to the first and the last residues in the alignments shown are presented in parentheses. Identical residues are in upper-case letters.

be involved in the formation of an ABC-type transporter (Finnie et al., 1997; Kröl and Skorupska, 1997).

The bt2 is another type of transporter having 59% homology with *Saccharomyces cerevisiae* Na^+-PO_4 cotransporter type III fragment (NPT2) of sodium-dependent phosphate transport integral membrane protein. The Na^+-PO_4 cotransporter (NPT2) play a role in active transport of phosphate into cells via sodium cotransport which belongs to the anion-cation symporters (ACS) family. These ACS are widely distributed in nature occurring in gram-negative and gram-positive bacteria, and in both animal and fungal eukaryotic kingdoms (Stephanie et al., 1998). From the sequence data obtained in this study, it is suspected that this cotransporter may be involved in the exchange of sodium and/or phosphate between the bacteroids and the cells of cv. Peking.

The ORF bt16 possesses 56% similarity to the *nif*E gene of *Azotobacter vinelandii* as against 47% similarity it shows to ABC transporter, Y4tE. This ORF bt16 and ORF bt7 may be directly involved in nitrogen fixation. The *nif*E gene (ORF bt16) lies within the Fix^+ segment (Fig. 3). The *nif*E is involved in the synthesis of nitrogenase molybdenum-iron (MoFe) cofactor. The MoFe component of nitrogenase (dinitrogenase) is directly involved in the final step of dinitrogen reduction to ammonia. Since NGR234 possesses a copy of this gene on its symbiotic plasmid (Freiberg et al., 1997), the dinitrogenase on the pBTBX may complements that of the wild-type NGR234 to establish an effective symbiosis with *G. max* cv. Peking.

The ORF bt7 encodes nitrogen fixation positive activator protein. This protein exhibits a similarity to the *nif*L of *Synechocystis* species, a nitrogen fixing cyanobacterium (Fig. 4). The *nif*L is a *nif*-specific regulatory gene which with *nif*A is involved in the global nitrogen regulatory (Ntr) system which controls *nif* genes expression in *Klebsiella pneumoniae* in response to environmental oxygen (Roelvinic and van den Bos, 1989; Fischer, 1994). Generally in rhizobial species aerobiosis directly interferes with NifA activity, the *nif*A is sensitive to oxygen *in vivo*. In *K. pneumoniae*, NifL is required for this control. The *nif*L also regulates the activity of NifA in response to nitrogen conditions in *K. pneumoniae*, a control mechanism that is not found in rhizobia (Fischer, 1994). It is assumed that, in addition to the activity of *nif*A in NGR234, the regulation of nitrogen fixation in the transconjugant, NGR234(pBTBX) is at least partially under the control of NifL. It is also suspected that the *nif*L homologue on the pBTBX renders the transconjugant NGR234(pBTBX) insensitive to nitrogen nutrient supply to the bacteroids by the cv. Peking cells, thereby enabling the bacterium to fix nitrogen effectively.

A *gcv*A (ORF bt8) homologue is located on the DNA. A comparison of the deduced amino acid sequence of ORF bt8 revealed homology to GCVA, a transcriptional

```
                                  **********
bt8 :          LPPLNALRMFESSARHLNFRVASEELGVTQGAVAQQVRALEQHLNTRLFDRRARGLELTEAGRRYMKPLQRA
GcvA :         LPPLNALRVFDAAARHLSFTRAAEELFVTQAAVSHQIKSLEDFLGLKLFRRRNRSLLLTEEGQSYFLDIKEI
LysR :       MAAVNLRHIEIFHAVMTAGSLTEAAHLLHTSQPTVSRELARFEKVIGLKLFERVRGRLH-PTVQGLRLFEEVQR
NodD :     MRFKGLDLNLLVALDALMTERKLTAAARSINLSQPAMSAAIGRLRAYFNDELFLMQQRRL-VPTPRAEALAPAVRE
AmpR :     MTRSYLPLNSLRAFEAAARHLSFTHAAIELNVTHSAISQHVKTLEQHLNCQLFVRVSRGLM-LTTEGENLLPVLND
                 354 4 3  4 4  3 333 33 4 5 4  34  33 4 4 34      44   3   55 4   4 5  4 3      3 3
Consensus :      PLN LR  F A  ARHLSFT AA EL VTQ AVS      LE   N   LF R   R L  L T      L P

bt8 :      LSLISDATQEL--KPATTVVTISVTPSFAAKWLIPRLGKFSDSNADTQVQVLAGNGLANFQSDGVDLAVRQTKPPF
GcvA :     FSQLTEATRKLQARSAKGALTVSLLPSFAIHWLVPRLSSFNSAYPGIDVRIQAVDRQEDKLADDVDVAIFYGRGNW
LysR :     SWYGLDRIVSAAESLREFRQGELSIACLPVFSQSFLPQLLQPFLAR-YPDVSLNIVPQESPLLEEWLSAQRHDLGL
NodD :     ALHIQLSVIAWDPLVPAESDRRFRIVLSDFMTLVFFEKVIKRVAREAPGVSFEL-LHVNDDPDERLRSGDLDFLIL
AmpR :     SFDRIAGMLDRFANH--RAQEKLKIGVVGTFATGVLFSQLEDFRRG-YPHIDLQLSTHNNRVDPAAEG---LDYTI
                       3  3        3       3             3          3   4
Consensus :            A  A        I       L             A          N   D

bt8 :      GPGLTGELLFPMRFIAVCSPTIVA-ESQIRTVEDLTRHVLLHDAHGM-WPVFLERAGVAADVRTLKSLRFSHSSLA
GcvA :     -PGLRVEKLYAEYLLPVCSPLLLTGEKPLKTPEDLAKHTLLHDASRRDWQTYTRQLGLN-HINVQQGPIFSHSAMV
LysR :     TETLHTPAGTERTELLSLDEVCV-LPPGHPLAVKKVLTPDDFQGENYISLSRTDSYRQLLDQLFTEHQVKRRMIVE
NodD :     PDQFMSATHPSAKLFEDKLVCVGCPSNQQLRGKLSLKRFMSMGHVAAMFGRTLKPSIEQW---LLLEHGFKRRVEI
AmpR :     RYGGGAWHGTEAEFLCHAPLAPLCTPDDIAASLHSPADILRFTLLRSYRRDEWTAWMQAAG------EHPPSPTHR
             3 3          3                                                  33 33  3     3
Consensus :  GL           L                                                  ADVL L     F

Bt8 :      IDAAVSGQGIALATEALVEDDLAAGRLCRPLDFAVTDELGFYIVHPRSPHKAGHVRSMRDWLLAQSHGSRQ(291)
GcvA :     LQAAIHGQGVALANNVMAQSEIEAGRLVCPFNDVLVSKNAFYLVCHDSQAELGKIAAFRQWILAKAAAEQE(305)
LysR :     THSAASVCAMVRAGVGISVVNPLTALDYAASGLVVRRFSIAVPFTVSLIRPHRPSSALVQA.....(300)
NodD :     VVPGFNSIPMLLQGTNRIATLPLLLVRHFEPTIPLQIVDHPLPPLSFTEALQWPLLHNSDP.....(300)
AmpR :     VMVFDSSVTMLEAAQAGVGIAIAPVDMFTHLLASERIVQPFATQIELGSYWLTRLQSRAETPA.....(291)
              3 3    3 33
Consensus :   A S    M LA
```

Figure 5. Multiple alignments of the amino acid sequence encoded by ORF bt8 with some LysR family proteins. GcvA, LysR, NodD and AmpR sequences are from *E. coli*, *E. coli*, Bradyrhizobia, *Enterobacter cloacae* respectively. Residue number corresponding to the last residue in the protein shown is presented in parenthesis. Numbers below the alignments represent the number of identical residues at a position where three or more matches are found. Helix-turn-helix residue position (asterisked) was identified based on the consensus motif (LTAAARALHLSQPAISRAIA; Henikoff et al., 1988) for LysR family proteins. Gaps introduced to make the best alignments are indicated with hyphens.

regulator. The GCVA functions as both a positive (by glycine) and negative (by purine, inosine) regulator of the glycine cleavage operon in *E. coli* (Wilson and Stauffer, 1994). It belongs to the LysR family. Transcriptional regulators, NodD and SyrM commonly found in rhizobia belong to the LysR family protein. This protein encoded by ORF bt8 has 291 amino acid residues containing a putative HTH DNA binding motif (Fig. 5). The hypothetical protein encoded by the ORF may therefore be involved in the transcriptional regulation of the genes upstream and/or downstream. Transcriptional regulators are host-range determinants in rhizobia-legume interaction.

The role played by the *gcv*A homologue located on the pBTBX in the interaction of NGR234(pBTBX) with cv. Peking is possibly similar to that of NodD and SyrM in regulating the biosynthesis of specific Nod factor required for host recognition and efficient nodulation of the soybean. This is substantiated by the fact that *nod*G (ORF bt9) homologue lies adjacent to the *gcv*A homologue (ORF bt8) on the pBTBX. The *nod*D genes control the first level of host specificity. The role played by *nod*G has been shown in certain rhizobial strains. Tn5 insertion mutation in *nod*G gene of *R. meliloti* resulted to a reduction in the number of nodules and a delay in nodule appearance (Horvath et al., 1986). Cloutier et al. (1997) reported reduced number and size of nodules on *Astragalus cicer* and *Onobrychis vicifolia* when inoculated with Tn5 *nod*G mutants of *R.* spp. strain N33, a Canadian high arctic rhizobial species. Therefore, ORF bt9, a *nod*G homologue is proposed to be involved in the synthesis of Nod factor in NGR234(pBTBX); possibly specifically in reduction reaction involved in the metabolism of fatty acids. The bt9 encoded a protein which might endowed the NGR234(pBTBX) with the ability to efficiently nodulate cv. Peking relative to the wild-type NGR234 which formed root-like structures. Another transcriptional regulator found on this DNA fragment has homology to GNTR transcriptional repressor of *Escherichia coli*. This protein

belongs to LacI family of transcriptional regulators. It negatively controls the induction of gluconate genes (*gnt*RKU) of the GntI system (Tong et al., 1996).

Methyl-accepting chemotaxis proteins (MCP) similar to the MCPs of *Rhodobacter capsulatus* and *Mycobacterium tuberculosis* are potentially encoded by ORFs bt14 and bt29. The MCPs are involved in motility. Motility confers a selective advantage on rhizobia in competition against non-motile strains (Ames and Bergman, 1981; Caetano-Anollès et al., 1988). The presence of *mcp* homologues on DNA in addition to the chemoreceptor genes in NGR234 may confer a selective advantage on the NGR234(pBTBX) transconjugant over other strains during its symbiotic interaction with soybean cv. Peking.

Recombinase, replication initiation protein, NADH-FMN oxidoreductase (encoded by *sna*C) and phosphoglycerate mutase are encoded by the DNA fragment. Recombinase (Cre) is part of a two components (Lox-Cre) of the bacteriophage P1 site-specific recombination system. It is essential for establishment of prophage and viral vegetative growth in a *rec*A$^-$ host (Sternberg et al., 1986). Phosphoglycerate mutase reversibly catalyzes the migration of phosphate ester from C3 to C2 in the conversion of 3-phosphoglycerate to 2-phosphoglycerate in glycolysis (Brock et al., 1984), a reaction which may contribute to the building up of sugar subunits during Nod factor synthesis important in the early stage of the symbiosis or in ATP generation coupled to the reduction of nitrogen.

This DNA sequence has shown that *R. fredii* USDA257 contains many genes encoding a large amount of theoretical proteins with potential roles assigned to them. This can be used to tackle problems of great importance in biotechnology, particularly where the manipulation of certain proteins may be of specific interest to legume yields. This is of special relevance in the case of *R.* species NGR234 that is promiscus, nodulating and/or fixing nitrogen on many legumes across the world. The use of this rhizobium in the development of inoculant will require the increase in its ability to nodulate and fix nitrogen on many more legumes; hence the significance of this sequencing to harness necessary genes.

Acknowledgement

We are grateful to the Nigeria and Switzerland Governments for funding this work.

REFERENCES

Ames, P. and Bergman, K. 1981. Competitive advantage provided by bacterial motility in the formation of nodules by *Rhizobium meliloti*. *Journal of Bacteriology* **148**: 728–729.

Balatti, P.A. and Pueppke, S.G. 1992. Identification of North American soybean lines that form nitrogen-fixing nodules with *Rhizobium fredii* USDA257. *Canadian Journal of Plant Science* **72**: 49–55.

Borofsky, M. 1995. GeneMark, version 2.3. Georgia Institute of Technology, School of Biology, Atlanta, GA, USA.

Brock, T.D., Smith, D.W., and Madigan, M.T. 1984. *Biology of Microorganisms*. Fourth edition. Prentice-Hall Int., Inc. pp. 108–113.

Caetano-Anollés, G., Wall, L.G., Micheli, A.T.D., Macchi, E.M., Bauer, W.D., and Favelukes, G. 1988. Role of motility and chemotaxis in efficiency of nodulation by *Rhizobium meliloti*. *Plant Physiology* **86**: 1228–1235.

Cloutier, J., Laberge, S., and Antoun, H. 1997. Sequence and mutational analysis of the 6.7-kb region containing *nod*AFEG genes of *Rhizobium* species strain N33: Evidence of DNA rearrangements. *Molecular Plant Microbe Interaction* **10:** 401–406.

Cooper, J.E. 2007. Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. *Journal of Applied Microbiology* **103:** 1355–1365.

Debellé, F., Moulin, L., Mangin, B., Dénarié, J., and Boivin, C. 2001. Nod genes and Nod signals and the evolution of the *Rhizobium*-legume symbiosis. *Acta Biochimica Polonica* **48:** 359–365.

Fellay, R., Rochepeau, P., Rélic, B., and Broughton, W.J. 1995b. Signals to and emanating from *Rhizobium* largely control symbiotic specificity. Histopathological, Biochemical, Genetic and Molecular Bases. Vol. 1: *Prokaryotes*. Singh, U.S. Singh, R.P., and Khomoto, K., eds. Pergamon Elsevier Science Ltd., Oxford. pp. 199–220.

Finnie, C., Hartley, N.M., Findlay, K.C., and Downie, J.A. 1997. The *Rhizobium leguminosarum prs*DE genes are required for secretion of several proteins, some of which influence nodulation, symbiotic nitrogen fixation and exopolysaccharide modification. *Molecular Microbiology* **25:** 135–146.

Fischer, H.M. 1994. Genetic regulation of nitrogen fixation in rhizobia. *Microbiology Review* **58:** 352–386.

Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A., and Perret, X. 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**: 394–401.

Freiberg, C., Perret, X., Broughton, W.J., and Rosenthal, A. 1996. Sequencing the 500kb GC-rich symbiotic replicon of *Rhizobium* species NGR234 using dye terminators and thermostable "Sequenase": a beginning. *Genome Research* **6:** 590–600.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**: 557–580.

Hanin, M., Jabbouri, S., Broughton, W.J., Fellay, R., and Quesada-Vincens, D. 1998. Molecular aspects of host-specific nodulation. In: *Plant-Microbe Interactions*. G.X. and Keen, N.T., eds. American Phytopathology Society, St. Paul, MN, USA. pp. 1–29.

Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. 1988. A large family of bacterial activator proteins. *Proceedings of the National Academy of Science, USA* **85**: 6602–6606.

Horvath, B., Kondorosi, E., John, M., Schmidt, J., Torok, I., Gyorgypal, Z., Barabas, I., Wieneke, U., Schell, J., and Kondorosi, A. 1986. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. *Cell* **46:** 335–343.

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.

Kröl, J. and Skorupska, A. 1997. Identification of genes in *Rhizobium leguminosarum* bv. *trifolii* whose products are

homologous to a family of ATP-binding proteins. *Microbiology UK* **143**: 1389–1394.

Lewin, A., Rosenberg, C., Meyer, Z.A.H., Wong, C.H., Nelson, L., Manen, J-F., Stanley, J., Dowling, D.N., Dénaire, J., and Broughton, W.J. 1987. Multiple host-specificity loci of the broad host-range *Rhizobium* species NGR234 selected using the widely compatible legume *Vigna unguiculata*. *Plant Molecular Biology* **8**: 447–459.

Paulsen, I.T., Beness, A.M., and Saier, M.H. Jr. 1997. Computer-based analyses of the protein constituents of transport systems catalyzing export of complex carbohydrates in bacteria. *Microbiology* **143:** 2685–2699.

Perret, X., Staehelin, C., and Broughton, W.J. 2000. Molecular basis of symbiotic promiscuity. *Molecular Biology Review* **64:** 180–201.

Pueppke, S.G. and Broughton, W.J. 1999. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular Plant Microbe Interaction* **12:** 293–318.

Rélic, B., Perret, X., Estrada-Garcia, M.T., Kopcinska, J., Golinowski, W., Krishnan, H.B., Pueppke, S.G., and Broughton, W.J. 1994. Nod factors of *Rhizobium* are a key to the legume door. *Molecular Microbiology* **13**: 171–178.

Roelvinic, P. and van den Bos, R.C. 1989. Regulation of nitrogen fixation in diazotrophs: the regulatory *nif*A gene and its characteristics. *Acta Botanica Neerlandica* **38**: 233–252.

Sambrook, J., Fritsch, E.F., and Maniatis, T.A. 1989. *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 136–147.

Shaw, L.J., Morris, P., and Hooker, J.E. 2006. Perception and modification of plant flavonoid signals by rhizosphere microorganisms. *Environmental Microbiology* **8**: 1867–1880.

Steinkellner, S., Lendzemo, V., Langer, I., Schweiger, P., Khaosaad, T., Toussaint, J.P., and Vierheilig, H. 2007. Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules* **12:** 1290–1306.

Stephanie, S., Pao, I., Paulsen, T., and Saier, M.H.Jr. 1998. Major facilitator superfamily. *Molecular Biological Review* **62**: 1–34.

Sternberg, N., Sauer, B., Hoess, R., and Abremski, K. 1986. Bacteriophage P1 *cre* gene and its regulatory region evidence for multiple promoters and for regulation by DNA methylation. *Journal of Molecular Biology* **187**: 197–212.

Tong, S., Porco, A., Isturiz, T., and Conway, T. 1996. Cloning and molecular genetic characterization of the *Escherichia coli gnt*R, *gnt*K and *gnt*U genes of Gnt1, the main system for gluconate metabolism. *Journal of Bacteriology* **178**: 3260–3269.

Trinick, M.J. 1980. Relationships amongst the fast-growing *Rhizobium* of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* species, *Acacia farnesiana* and *Sesbaenia grandiflora* and their affinities with other *Rhizobium* groups. *Journal of Applied Bacteriology* **49**: 39–53.

Wilson, R.L. and Stauffer, G.V. 1994. DNA sequence and characterization of GcvA, a LysR family regulatory protein for the *Escherichia coli* glycine cleavage enzyme system. *Journal of Bacteriology* **176**: 2862–2868.