Molecular Microbiology (1999) **32**(2), 415–425

High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234

Xavier Perret,[1†] Christoph Freiberg,[2†] André Rosenthal,[2] William J. Broughton[1*] and Rémy Fellay[1]
[1]Laboratoire de Biologie Moléculaire des Plantes Supérieures, University of Geneva, 1 chemin de l'Impératrice, 1292 Chambésy, Geneva, Switzerland.
[2]Institut für Molekulare Biotechnologie, Abteilung Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany.

Summary

Most of the bacterial genes involved in nodulation of legumes (*nod*, *nol* and *noe*) as well as nitrogen fixation (*nif* and *fix*) are carried on pNGR234*a*, the 536 kb symbiotic plasmid (pSym) of the broad-host-range *Rhizobium* sp. NGR234. Putative transcription regulators comprise 24 of the predicted 416 open reading frames (ORFs) contained on this replicon. Computational analyses identified 19 *nod* boxes and 16 conserved NifA-σ^{54} regulatory sequences, which are thought to co-ordinate the expression of nodulation and nitrogen fixation genes respectively. To analyse transcription of all putative ORFs, the nucleotide sequence of pNGR234*a* was divided into 441 segments designed to represent all coding and intergenic regions. Each of these segments was amplified by polymerase chain reactions, transferred to filters and probed with radioactively labelled RNA. RNA was extracted from bacterial cultures grown under various experimental conditions, as well as from bacteroids of determinate and indeterminate nodules. Generally, genes involved in the synthesis of Nod factors (e.g. the three *hsn* loci) were induced rapidly after the addition of flavonoids, whereas others thought to act within the plant (e.g. those encoding the type III secretion system) responded more slowly. Many insertion (IS) and transposon (Tn)-like sequences were expressed strongly under all conditions tested, while a number of loci other than those known to encode *nod*, *noe*, *nol*, *nif* and *fix* genes were also transcribed in nodules. Many more diverse transcripts were found in bacteroids of determinate as opposed to indeterminate nodules.

Received 26 October, 1998; revised 21 January, 1999; accepted 26 January, 1999. †These authors contributed equally to this study. *For correspondence. E-mail broughtw@sc2a.unige.ch; Tel. (+41) 22 906 1740; Fax (+41) 22 906 1741.

Introduction

Soil bacteria of the genera *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* (collectively known as rhizobia) are capable of forming symbiotic associations with leguminous plants. Under conditions of limiting soil nitrogen, specialized root structures called nodules may develop. In compatible interactions, rhizobia are released into the nodules, where they form nitrogen-fixing bacteroids which exchange ammonia for nutrients synthesized by the host plant. Symbiotic specificity varies from rhizobia that are devoted to one or a few legumes to *Rhizobium* sp. NGR234, which nodulates more than 112 genera of legumes (and the non-legume *Parasponia andersonii*) (Trinick, 1980; Lewin *et al*., 1987; Pueppke and Broughton, 1999). Transconjugants of *Agrobacterium tumefaciens* cured of its Ti plasmid, but harbouring the symbiotic plasmid of NGR234 (pNGR234*a*), are able to nodulate *Vigna unguiculata* (Broughton *et al*., 1984). This implies that most symbiotic loci are carried on pNGR234*a*, although homologues of *nodEG* and *nodPQ* were later found on the chromosome (Perret *et al*., 1991). Further analysis revealed that the molecular determinants of promiscuity are clustered in three, dispersed (Perret *et al*., 1991) *hsn* (host specificity of nodulation) loci (Broughton *et al*., 1986; Lewin *et al*., 1987). In contrast to *nod* genes, most *nif* and *fix* homologues are grouped into a single 55 kb cluster (*fixU* to y4xE), which has been spared disruption by insertion (IS) or transposon-like sequences (Freiberg *et al*., 1997). So far, only four of the 24 putative regulators of transcription encoded by pNGR234*a* have been the subject of detailed investigations: *nodD*1 (Relić *et al*., 1993a), *nodD*2 (Fellay *et al*., 1998), *syrM*1 (Hanin *et al*., 1998) and y4sM (Perret *et al*., 1994).

Regulation of bacterial and plant symbiotic genes is mediated by the exchange of molecular signals between the two symbionts. Among the many compounds present in root exudates, flavonoids are potent activators of *nod* gene transcription (Fellay *et al*., 1995a). Expression of nodulation genes is mediated by NodD proteins, which belong to the LysR family of prokaryotic transcriptional regulators. Although NodD binds to specific promoter sequences called *nod* boxes even in the absence of an inducer,

the presence of flavonoids reinforces this interaction and is required for transcriptional activation (Goethals *et al*., 1992; Fisher and Long, 1993). In turn, most *nod* genes are involved in elaboration of another class of bacterial signal molecules, the Nod factors. These lipochito-oligosaccharides permit bacteria to enter the root and induce the formation of nodule meristems (Relić *et al*., 1993a and b; 1994; Dénarié *et al*., 1996; D'Haeze *et al*., 1998; Hanin *et al*., 1999). Precise modulation of *nod* gene expression is required in effective symbioses. For instance, NolR, which is present in many rhizobia (Kiss *et al*., 1998), is needed for optimal nodulation of *Medicago sativa* by co-ordinating the repression of *nodABC* in *R. meliloti* (Kondorosi *et al*., 1989; Cren *et al*., 1995). Similarly, down-regulation of the *nodABCIJnolOnoeI* operon, which occurs after the initial flavonoid induction, is abolished in NGRΩ*nodD*2 and results in modification of the symbiotic properties of the mutant strain (Fellay *et al*., 1998).

During the later stages of the symbiosis, the expression of many bacterial genes is co-ordinated with nodule morphogenesis via decreasing oxygen pressure in infected nodule cells (Vasse *et al*., 1990; Fisher, 1994). Although the regulation of *nif* and *fix* genes differs between *A. caulinodans*, *B. japonicum* and *R. meliloti*, all require FixL and FixJ (an oxygen-dependent, two-component regulatory system), FixK (a transcriptional regulator whose expression is activated by FixJ), as well as NifA and RpoN. So far, homologues of *fixLJ* and *fixK* have not been identified in NGR234, but both the pSym-borne *nifA* and the chromosomal *rpoN* genes are required for the formation of fully effective nodules (i.e. Fix^+) (van Slooten *et al*., 1990; R. Fellay, unpublished).

Of the several ways to assay the symbiotic function of the plasmid-borne genes, mutagenesis of 416 open reading frames (ORFs) followed by screening for altered phenotypes on 112 genera of legumes is clearly impractical. Instead, we developed methods of identifying those genes that are differentially expressed during the symbiosis. Previous transcriptional analyses of entire symbiotic replicons were based on the physical maps of pNGR234*a* (Perret *et al*., 1991) and the pSym of *R. etli* strain CFN42 (Girard *et al*., 1991). DNA of overlapping cosmids was digested with restriction enzymes and probed with radioactively labelled RNA of bacteroids or rhizobia grown under various environmental conditions (Fellay *et al*., 1995b; Girard *et al*., 1996). Unfortunately, it was not possible using these methods to determine which of the genes found on a large hybridizing restriction fragment was actively transcribed. To obtain higher resolution, pNGR234*a* was divided into 441 segments selected to represent all ORFs as well as their intergenic regions. After polymerase chain reaction (PCR) amplification, these segments were separated on agarose gels, transferred to membranes and probed with radioactively labelled RNA.

Results

Transcription of some nodulation loci, such as the *nodABC* genes, can be detected a few minutes after the addition of flavonoids to the growth medium. Time course experiments showed that the expression of most *nod* genes reaches a maximum after 1 h of induction with flavonoids and, as in the cases of the *nodABCIJnolOnoeI* and *nodSU* transcripts, was no longer detected 24 h later (Fellay *et al*., 1995b). To compare our results with those of earlier studies based on competitive RNA hybridizations, liquid cultures of NGR234 were grown in the presence of daidzein for 1 h and 24 h. Genes expressed in nodules were identified by labelling RNA extracted from bacteroids purified from plants that form determinate (*Vigna unguiculata*) as well as indeterminate nodules (*Cajanus cajan*).

Reproducibility and sensitivity of the technique

Background signal artefacts were excluded from the analysis by comparing the positions of PCR products on filters with those of hybridization signals on autoradiographs (Fig. 1). To facilitate interpretation of the data, hybridization intensities were classified into seven categories ranging from none (0) to maximum (M) (see Fig. 1; Fig. 2). In the absence of a reliable internal standard (e.g. a gene whose transcription remains constant in free-living cells as well as in bacteroids), the autoradiographs were exposed until the signal of some loci reached maximum intensity (see Fig. 1, panel sKL to ft1). In this way and in all experiments, several DNA fragments produced signals of similar intensities. Reproducibility of the method was assessed in various ways. Most experiments, including controls in which the cells were grown in liquid cultures without flavonoids, were repeated twice. No significant differences in the hybridization patterns of duplicate experiments were observed, and the results are consistent with previously published hybridization data (Fellay *et al*., 1995b; Freiberg *et al*., 1997; Rochepeau *et al*., 1997; Viprey *et al*., 1998). Furthermore, observations made using β-galactosidase assays (Berck *et al*., 1999; Fellay *et al*., 1998; Hanin *et al*., 1997) and β-glucuronidase promoter fusion studies (A. Del Greco and X. Perret, unpublished) fully supported the data shown in Fig. 2.

Although expression of 169 ORFs was not detected, radioactively labelled RNA hybridized to about 60% (247) of the 416 genes and gene fragments (Fig. 2). Undoubtedly, many factors, including the half-life of the individual mRNAs, affect signal intensity. Also, actively transcribed genes that are duplicated on another replicon add to the hybridization signals. An example of this kind of interference is obvious in the case of *dctA*1, which is part of an operon that begins with *nifQ* and is controlled by a NifA-σ^{54}-dependent promoter. This locus is expressed in nodules (Fig. 2, segment

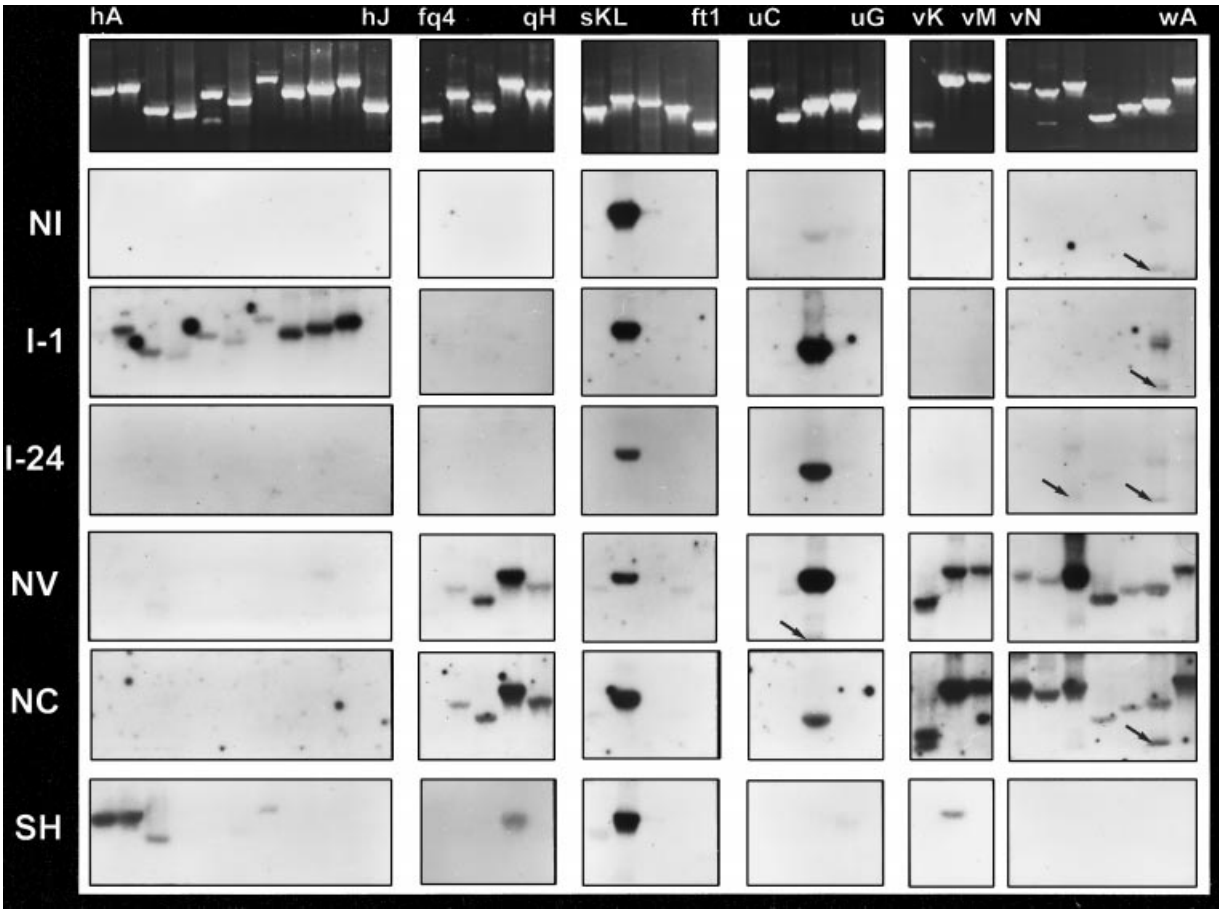
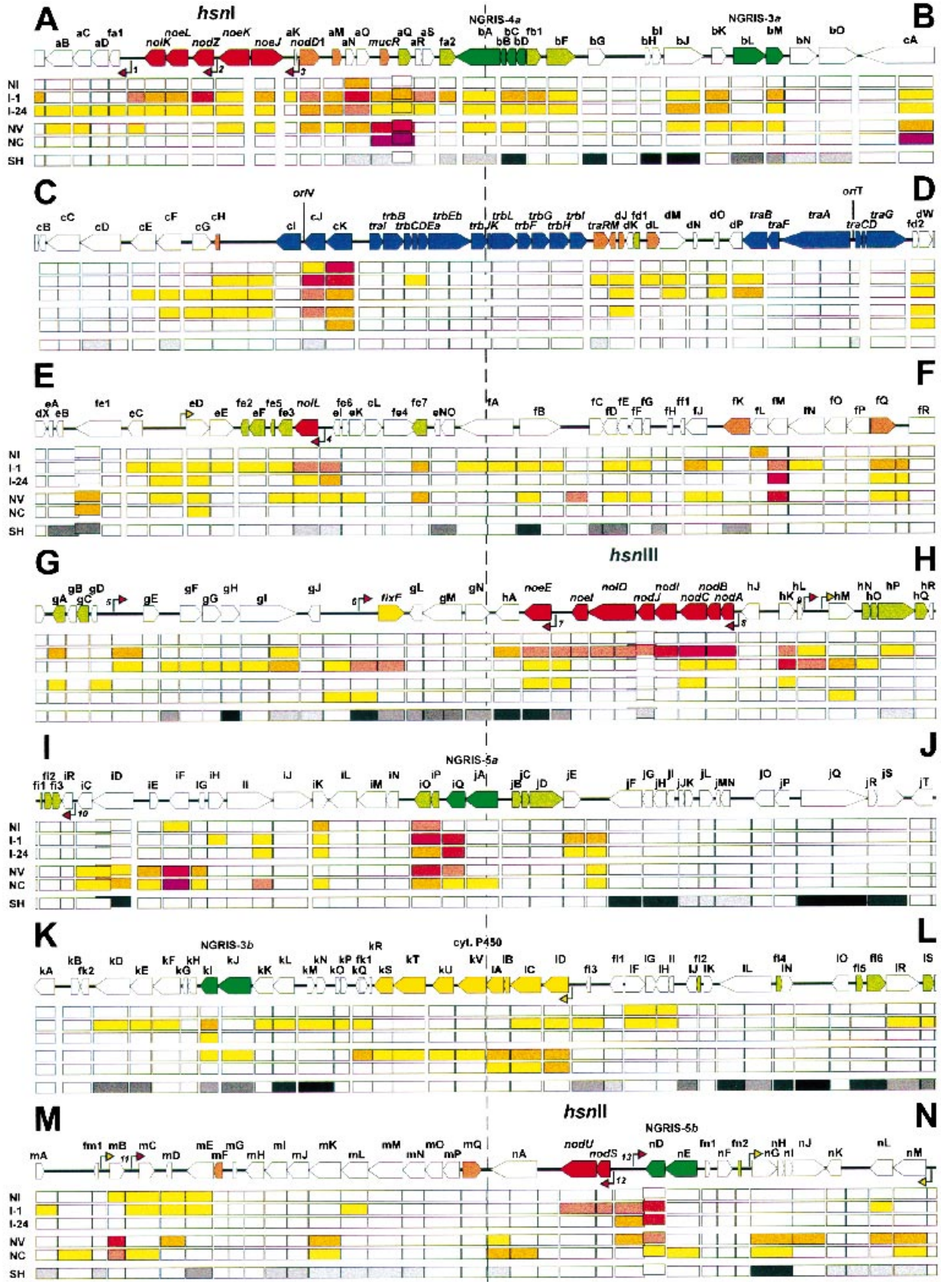


Fig. 1. Examples of the transcription analyses performed on pNGR234*a*.The upper panels correspond to ethidium bromide-stained gels of PCR products separated on 1% agarose gels. Letters above each column represent the first and last PCR fragment shown in a panel. The lower panels correspond to hybridization signals in: NI, NGR234 cells grown at 27°C in liquid *Rhizobium* minimal medium supplemented with succinate (RMM); I-1, cells grown in RMM followed by a 1 h induction with 2×10^{-7} M daidzein; I-24, RMM-grown rhizobia collected 1 day after induction with daidzein; NV, bacteroids purified from Fix^+ nodules of *V. unguiculata* inoculated with NGR234; NC, bacteroids purified from Fix^+ nodules of *Cajanus cajan* inoculated with NGR234; and SH, fragments probed with NGR234 sequences purified by subtractive DNA hybridization against genomic DNA of *R. fredii* strain USDA257. Intensities of hybridization signals ranged from none (0) to maximum (M). For example, in row NC, panel fq4 to qH, values of 0, 3, 4, 5 and 4 were attributed to signals from left to right respectively. In contrast, the signal in the second lane of the adjacent panel sKL to ft1 (row NC) corresponds to the maximum intensity (M). Non-specific hybridization to PCR by-products are marked with arrows.

V) but, after induction with daidzein, the weak signals detected with the *dctA*1 fragment are probably caused by transcripts from the chromosomal copy (*dctA*2) (Fig. 2, I-1 and I-24). Fortunately, non-specific cross-hybridizations are restricted to nearly identical DNA sequences. For example, ORFs y4aM and y4wC encode putative transcriptional regulators of 69% identity and 89% similarity but y4aM (Fig. 2, segment A, I-1 and I-24) is induced early and expressed in *V. unguiculata* nodules, while y4wC is induced late and is expressed in both determinate and indeterminate nodules (Fig. 2, segment W, NV and NC). It thus seems as though these techniques can discriminate between transcripts that show 82% identity over more than half the length of the genes.

*Expression of pNGR234*a *genes in non-induced conditions*

Many of the 22 ORFs expressed in free-living cells (Fig. 2, NI) are also transcribed under all conditions. Among these are *repC* (y4cK), which is involved in the replication of pNGR234*a*, y4xL, a protein excreted by the type III secretion system (TTSS)(Viprey *et al*., 1998), a protein of 272 amino acids of unknown function (y4pM) as well as products of IS-like sequences (y4iO and y4iP; y4pG and


A
hsnI
B
NGRIS-4a
NGRIS-3a
noeL
noeK
noeJ
nodD1
mucR
C
D
orfV
orfT
trbB
trbEb
trbL
trbG
trbI
traRM
traA
traG
E
F
nolL
G
hsnIII
H
noeE
noeI
nolO
nodI
nodJ
nodB
nodA
I
J
NGRIS-5a
K
L
NGRIS-3b
cyt. P450
M
hsnII
N
NGRIS-5b
nodU
nodS
NI
I-1
I-24
NV
NC
SH

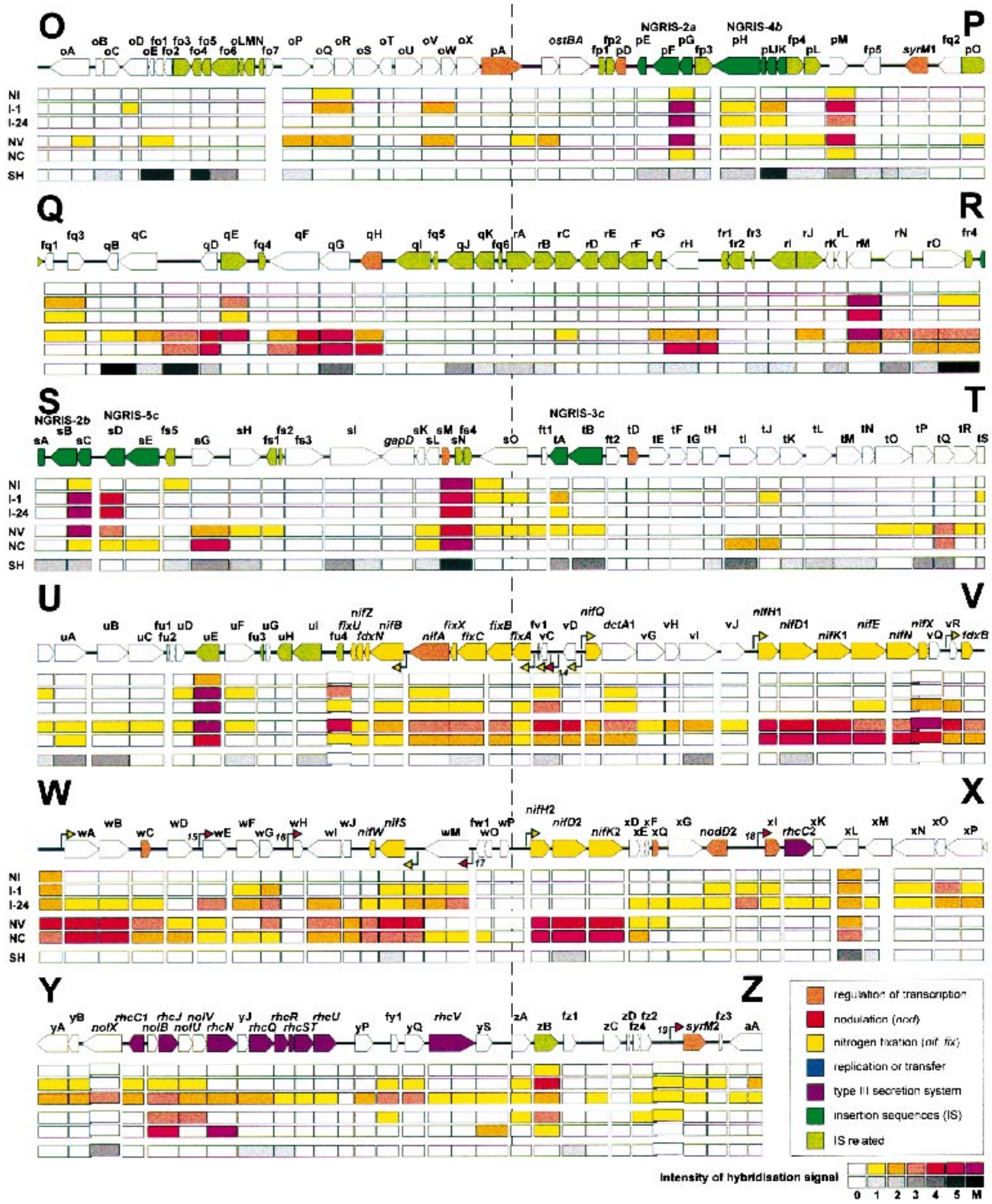


Fig. 2. Transcription map of pNGR234*a*. Genes (e.g. bF in segment B) and gene fragments (e.g. fb1 in segment B) are represented by arrows showing the direction of transcription and coloured depending on their predicted function. The position and orientation of putative symbiotic promoters is shown with small coloured arrows: *nod* boxes are marked in red and numbered as in Table 1, NifA-σ^{54} sequences are shown in yellow. Size and position of PCR-amplified fragments corresponding to gene(s) or intergenic regions are given directly below the genetic map. Overlapping fragments are displaced with respect to adjacent products. Colour-coded intensities of expression range from white (0 = no signal) to purple (M = maximum) in seven steps. RNA probes were prepared from: NI, NGR234 cells grown at 27°C in liquid *Rhizobium* minimal medium supplemented with succinate (RMM); I-1, cells grown in RMM followed by a 1 h induction with 2×10^{-7} M daidzein; I-24, RMM-grown rhizobia collected 1 day after induction with daidzein; NV, bacteroids purified from Fix$^+$ nodules of *V. unguiculata* inoculated with NGR234; NC, bacteroids purified from Fix$^+$ nodules of *Cajanus cajan* inoculated with NGR234; and SH, the 441 fragments were hybridized against pooled sequences made specific by subtractive DNA hybridization against genomic DNA of *R. fredii* strain USDA257. Hybridization intensities are shown in shades of grey.

y4sC of NGRIS-2*a* and -2*b* respectively; fs5; y4sN and fs4; y4uE). Other loci, such as y4lF, y4lG, y4lH and y4lL, are only expressed in non-induced cultures (as well as in cells exposed to a flavonoid for 1 h), suggesting that they might contribute to growth under free-living conditions.

Flavonoid induction of transcription

The addition of flavonoids to the growth medium dramatically changed the patterns of gene expression (Fig. 2, I-1 and I-24). Daidzein enhanced the transcription of approximately 147 previously silent ORFs. Of these, 14 belong to three classes of insertion sequences (NGRIS-3, IS-4 and IS-5). As these ISs are present in multiple and identical copies in pNGR234*a* as well as in other replicons, it is clearly not possible to determine which of these elements carry the activated gene(s).

Except for several ORFs linked to IS elements, most of the strongly induced genes are found in the three *hsn* loci and the cluster encoding the type III secretion system (TTSS). With the exception of y4xL, which is thought to be constitutively expressed (Viprey *et al*., 1998), the expression of all ORFs between y4xI and y4yS occurred 24 h after induction with daidzein. In contrast, expression of most *nod* genes (marked in red in Fig. 2) reached a maximum after 60 min of induction and declined later. This is particularly clear in the case of the *nodABCIJnolOnoeI* operon (Fig. 1, I-1 and I-24, first column), transcription of which is repressed in a *nodD2*-dependent manner 24 h after flavonoid induction (Fellay *et al*., 1995a; 1998). Interestingly, levels of *nodD*1 transcripts were increased significantly after flavonoid induction (Fig. 2; I-1, segment A). A similar observation was made in *R. fredii* USDA191, a strain closely related to *R. fredii* USDA257. As in NGR234, *nodD*1 of USDA191 and USDA257 are immediately downstream of a conserved *nod* box-like sequence (Table 1 and Fig. 2; *nod* box no. 3). Although present in the reverse orientation, NodD1 and NodD2 bind to this motif and modulate the expression of *nodD*1 of USDA191 (Machado *et al*., 1998).

Computational analyses of the pNGR234*a* sequence revealed 19 *nod* box-like regulatory sequences, at least five of which (*nod* box nos. 2, 4, 7, 8 and 12; Table 1; Fig. 2) regulate the expression of known *nod* loci. With the exception of *nod* boxes nos. 10 and 16, the transcription patterns of downstream ORFs suggest that most of these regulatory elements are functional. The apparent lack of promoter activity of *nod* boxes 10 and 16 is probably not the result of accumulated mutations, however, as major differences between the DNA sequences of these elements and consensus *nod* boxes (van Rhijn and Vanderleyden, 1995) were not found (Table 1). In contrast, *nod* box no. 13 appears to have retained regulatory activity, despite the absence of properly oriented downstream genes. That one of the three identical, pSym-borne copies of NGRIS-5 is adjacent to *nod* box no. 13 implies that a gene was probably separated from its promoter by rearrangements in the genome.

Transcription of y4gF, y4gG, y4gH and y4gI, which are probably involved in the biosynthesis of rhamnose, was only detected 24 h after induction with daidzein. As with some other flavonoid-inducible loci (e.g. y4fA, y4fB, y4iH, y4jE), their regulation cannot be directly associated with *nod* boxes. This points to the existence of alternative or indirect regulatory pathways, some of which nonetheless require a functional NodD1. For instance, some genes of the TTSS locus are thought to be under the control of y4xI, a *nod* box-controlled regulator of transcription (Viprey *et al*., 1998).

Gene expression in nodules

Important changes in gene expression accompany the differentiation of rhizobia into nitrogen-fixing bacteroids (Fig. 2, NV and NC). Principally, these involve repression of the majority of flavonoid-inducible transcripts and induction of those required for effective nitrogen fixation. Because of the marked sensitivity of nitrogenase to oxygen and the high energy requirements of nitrogen fixation, the expression of *nif* and *fix* loci is tightly controlled (for review, see Fischer, 1994). Unlike *nod* genes, which are dispersed over the complete pSym, *nif* and *fix* genes form a 55 kb cluster (*fixU* to y4xE) that contains 10 NifA-σ^{54}-dependent promoters. Transcriptional analyses suggest that many more loci are specifically expressed in bacteroids, however. Some, such as the cytochrome P450 operon (y4kR to 'y4lD), are under the control of one of the six other NifA-σ^{54}-like sequences. Two opposing operons, y4nG to y4nJ and y4nL to y4nM, which possibly encode functions linked to sugar metabolism, are also regulated in a similar manner.

As mentioned above, flavonoid-inducible *nod* boxes do not necessarily precede *nod* genes. Similarly, a number of NifA-σ^{54}-independent genes were actively expressed in nodules. These include two co-transcribed genes, y4qG and y4qF, which encode a class III pyridoxal-phosphate-dependent aminotransferase and a peptidase of the S9A family respectively. Based on the intensity of hybridization signals, both ORFs are expressed at levels as high as those of the nitrogenase structural genes (see Fig. 1, NV, NC, panels fq4 to qH and vK to vM), suggesting an important role in the functioning of determinate and indeterminate nodules. Although y4sO is clearly homologous to y4qF (70% identity and 92% similarity at the amino acid level), y4sO is apparently not transcribed in nodules.

Table 1. List of predicted *nod* boxes contained in pNGR234*a*.

							Induction[a]	
nod box	Position	Orientation	Sequence	Mismatch[b]	Distance[c]	Gene	1 h	24 h
1	4538	−	A ATC TACCGTATA GAT TTATGTC ATT GAAAAAATG GAT TTCAATAAAG C	11	503	fa1	0	1
2	8503	−	C ATC CAATATGTG GAT GGTTGCC ATT CAAAATATA AAT TTTACTGATG T	8	87	*nodZ*	4	1
3	12 346	−	A ATC CACAGTGCG GAT GGTTGTC ATC CGAACGATC AAT TATGCAAATC G	7	\[d]	\[d]	1	0
4	97 494	−	A ATC CATCATGTA AAT GCGGTTC ATC CAAACAATT GAT TGCACCAATT T	6	276	*nolL*	3	1
5	129 639	+	A TTC CATGGTTTG AAT AGGCCCC ATA CAAAAAATG GAT TTTCCCAGTC T	10	1358	y4gE	2	1
6	141 112	+	T ATC CATAGTGCT GAT GAGTGGC ATG CAAACATTC AAT TTTTCCAGCG C	8	890	*fixF*	0	3
7	150 304	−	C ATC CAAAGCATG GAT GTGTTGC ATT GGCAGAAAC GAT TTGACCGGGC T	11	201	*noeE*	3	1
8	158 844	−	T ATC CACAGGATG AAT GGCTATA ATC CAAACAATC AAT TTTACGGATC C	4	235	*nodA*	5	1
9	161 915	+	T ATC CAAAGTGCG GAT AATCGTG ATG CAAAATATT TAT TTTACTGGCC T	11	1103	y4hM	1	3
10	169 857	−	A TTC CATGGTGTG GAT AGGCCCC ATA CAGAAAATC GAT TTTCCCAATC T	7	117	y4iR	0	0
11	256 458	+	C ATC TACGCAGTG GAT CTTTGCT ATC CACACAACC GTT TTGGCAATCT T	12	329	y4mC	1	0
12	278 971	−	C TTC CATAACCTG GAT GTTTGGC ATC ACAAAAATC GAT TTTACGATTT T	8	153	*nodS*	3	0
13	279 846	+	C ATC CATGGCGTG GAT CTTTGCC ATC CAAATAAAC AGT TATCGCAATT T	8	\[d]	\[d]	3	2
14	443 125	−	G ATC CATTGAGCT AAT GCGTGCC ATC GAGATAATC GAT TTCACCGAGA T	10	465	y4vC	1	3
15	469 175	+	A ATT CATATCATC AAT GGCTGCG ATG AAAACAAAT GAT TTTACTGAAT A	12	112	y4wE	0	3
16	473 083	+	C TTC CATCTCCTC AAT AGCCTCC ATC CAAATAAGC GAT TTTACATATC T	9	236	y4wH	0	0
17	481 277	−	T ATC CATAGACGC TAT GAACAGA ATT TAGAAAATC AAT TTTTGAGGAG T	16	117	y4wM	1	3
18	493 985	+	C ATC CATGATGCA GAT GCCTTCG ATC CGAAGAATC GAT TTGCTCAATC A	6	288	y4xI	1	1
19	532 063	+	C ATC CACGGTATG GAT GATAGCT ATG CAAAAACTC GAT TTTTCCATTC T	5	589	*syrM2*	1	1
Consensus[e]			Y ATC CAYNNYRYR GAT GNNNNYN ATC NAAACAATC RAT TTTACCAATC Y					

Sequences of *nod* boxes found on pNGR234*a* are centred around the given positions (Position). Depending on whether they are found on the plus or minus DNA strand, their orientation (Orientation) is marked as + or − respectively. *nod* boxes are numbered from 1 to 19 by order of appearance, starting at base one of segment A (as in Freiberg *et al*., 1997).

a. Detection of transcriptional activity 1 h (1 h) and 24 h (24 h) after induction with flavonoids (expression levels as in Fig. 2).

b. Number of mismatches compared with the consensus sequence.

c. Distance to the next downstream gene present in proper orientation (Gene).

d. No gene was found in the proper orientation downstream of the *nod* box.

e. *nod* box consensus sequence as in van Rhijn and Vanderleyden (1995), with R for either A or G; Y replacing U, T or C; N is for any base.

Differential gene expression in determinate and indeterminate nodules

Significant differences in gene expression were found when comparing the transcription patterns in bacteroids of *V. unguiculata* with those of *C. cajan* (Fig. 2, NV and NC). Except for genes directly involved in nitrogen fixation (e.g. *nif*, *fix*, and cytochrome P450 loci), which are transcribed in both nodule types, more than 20 ORFs were only expressed in determinate nodules. Among these were *nodD*1, y4aN (homologous to *syrB* of *R. meliloti*), y4aO, a homologue of *A. rhizogenes repB* (y4cJ), y4oP to y4oR (coding for components of an ABC transporter system), two pioneer ORFs (y4oV and y4oW), *otsB* (involved in trehalose synthesis), two co-transcribed genes (y4qC and y4qB), several IS-like sequences (fe7, y4qE, y4rG and y4rJ), y4rN (coding for a transmembrane protein homologous to an antibiotic resistance determinant of *Mycobacterium smegmatis*), a transcription regulator (y4xQ), as well as *nolU* and *nolV* of the TTSS cluster. In contrast, fewer genes were clearly expressed in *C. cajan* nodules. Among them are an oxidoreductase (y4hM), peptidases (y4nA and y4tI), a transposase (y4nE), a threonine dehydratase (y4tJ), an ABC transporter binding protein (y4wM) and RhcN, whereas y4mC and y4wO did not have any homologous database entries.

Subtractive DNA hybridization

To identify DNA sequences of NGR234 that are not found in the genome of the closely related strain *R. fredii* USDA257, a pool of *Sau*3AI fragments was purified by subtractive DNA hybridization (Perret *et al*., 1994). Probing of the filter set with these radioactively labelled sequences showed that many of the 441 fragments hybridized (Fig. 2, SH). Signals found with the *nifD* and *nolX* fragments (Fig. 2, SH, segments X and Y) raise questions about the ability of the technique to eliminate every common sequence, especially as the *nolX* fragment of NGR234 is 98% identical to that of the PCR fragment from USDA257.

Nonetheless, many of the 150 hybridizing fragments (i.e. those absent from USDA257) were expected from the known differences between the two genomes. Among these were *nolL* and *noeE* (Fig. 1, SH, panel hA to hJ), which are responsible for the acetylation and sulphatation of NodNGR factors, respectively (Hanin *et al*., 1997; Berck *et al*., 1999), as well as a series of ORFs that comprise *fixF* to y4hA (Fig. 2, SH, segment G and H). Others include the NGRRS-1 locus, of which the clustered NGRIS-2, NGRIS-4 and NGRIS-10 elements are absent from USDA257 (Perret *et al*., 1997). In fact, over 20% of the fragments hybridizing to subtracted sequences correspond to IS- and Tn-like elements, suggesting that a significant proportion of the

differences between the NGR234 and USDA257 genomes comprises mobile elements.

Downstream of *hsn*II, a cluster of genes unique to NGR234 seems to be involved in the transport of sugars and other compounds (y4mI, y4mJ and y4mK), whereas another set of genes (y4tI to y4tN) encodes various classes of enzymes. Other unique genes are possibly under the control of NifA-σ^{54} promoters. These include y4nG to y4nI, which encode a putative nucleotide sugar epimerase, as well as two transmembrane proteins. Genes encoding other general metabolic functions seem to occur only on pNGR234*a*, including an aminotransferase (y4uB) and y4uA, an enzyme involved in the biosynthesis of cell wall components.

Discussion

Although sequences unique to NGR234 represent a sizeable fraction of the pSym, most genes found on pNGR234*a* are also present in USDA257. Obviously, unique *Sau*3AI fragments are not clustered in a single locus. Rather, they form dispersed islands, highlighting the dynamic structure of this plasmid. Large repeats, IS and Tn sequences are possibly involved in this phenomenon. The transcriptional activation of ORFs belonging to IS elements during flavonoid induction or in bacteroids suggests that transposition may occur at higher frequencies during symbiosis. Apart from the known differences in *nod* genes (Pueppke and Broughton, 1999) and IS elements (Perret *et al*., 1997), USDA257 also lacks various enzymes, some of which are expressed in nodules (e.g. y4qB to y4qD). Numerous fragments that hybridize to sequences unique to NGR234 also carry genes of unknown function, many of which are not expressed under the conditions tested.

Similarly, a large number of predicted ORFs apparently remain silent. Perhaps the methods used failed to detect very low-level gene expression, but the conditions tested here represent only a small subset of those encountered by NGR234 during its life cycle. Nonetheless, flavonoid-inducible and bacteroid-specific genes make up 60% of the 441 ORFs and gene fragments encoded by pNGR234*a*. Several large intergenic regions, such as fa1-*nolK*, y4cH-y4cI, y4eC-y4eD, y4qC-y4qD and *fdxB*-y4wA, encode transcripts induced by flavonoids that may or may not be found in bacteroids. Analysis of these regions using a new and more sensitive GeneMark matrix compiled from the genes of pNGR234*a* failed to detect ORFs however. This is possibly because of genes that: (i) use very distinct codons; (ii) have ORFs that are too short to be detected using current programmes; or (iii) encode symbiotically active but non-translated RNA species (this has been described in *B. japonicum*; Ebeling *et al*., 1991).

In general, the timing of expression of individual genes correlates well with the presence of *nod* box or NifA-σ^{54}-regulatory sequences. Genes dependent on *nod* boxes are flavonoid inducible and expressed early in the symbiosis, whereas NifA-σ^{54}-dependent loci are activated in bacteroids. Nevertheless, y4vC is apparently controlled by both (*nod* box no. 14 and a NifA-σ^{54}-regulatory sequence). Remarkably, Nod factor biosynthetic genes are not organized into a singular regulon. Although most *nod* gene promoters contain *nod* boxes, expression of the *nodABCIJnolOnoeI* and *nodSU* operons is abolished 24 h after induction with daidzein, whereas transcription of *noeE* and *nolL* can still be detected in *V. unguiculata* nodules. Obviously, several parallel and probably interdependent regulatory pathways control the expression of symbiotically active genes. Daidzein induction of expression of the transcriptional regulators *nodD*1, y4aM, y4fQ, *nodD*2, y4xI and *syrM*2 (the last two of which are under the control of active *nod* boxes), and the increased transcript levels of y4qH and y4wC found in bacteroids, highlights this complexity.

Unfortunately, the charged membranes used for hybridization could only be reused once, severely limiting the number of analyses that could be performed. Thus, we were unable to analyse transcription patterns in several different determinate and indeterminate nodules. Nevertheless, the available data point to important differences between bacteroids of *V. unguiculata* and *C. cajan* nodules. Transcription of a number of *nod* genes, a gene involved in plasmid replication, ORFs encoding transporters of sugars as well as amino acids/peptides, a regulator of transcription of various biosynthetic and respiratory enzymes suggest that *V. unguiculata* bacteroids have a higher overall metabolic activity than those found in *C. cajan* nodules. Although fewer ORFs are expressed in *C. cajan* nodules, some are nonetheless specific to this plant, suggesting that bacteroid metabolism varies with the host.

Despite its current limitations, our system of transcriptional analyses is a solution to the problem of quickly and accurately delimiting important symbiotic loci and their regulatory elements. Separation of PCR products on gels before Southern blotting followed by careful recording of the positions of correct fragments on membranes helped to eliminate false-positive signals (marked with arrows in Fig. 1). A necessary refinement to the technique would be to produce membranes (or arrays) that could be probed many times. This would permit analyses of regulatory pathways by performing expression studies with RNA extracted from, for example, *nodD*1, *nodD*2, *syrM*1 and *nifA* mutants. Analyses could also be made using RNA extracted from NGR234 cells grown under low oxygen or high salinity, or induced with root exudates rather than purified flavonoids. Yet the present technique will help to dissect symbiotic molecular genetics. All 19 *nod* boxes will be cloned,

fused to an appropriate reporter gene and their activity in the presence of different flavonoids and/or root extracts assayed. Genes such as y4aN, y4aO, y4qF and y4qG, as well as the regulators of transcription y4aM, *mucR* and y4qH can be mutated and their effect on symbiosis in determinate versus indeterminate nodules assayed. Loci y4gF to y4gI, which encode enzymes involved in rhamnose synthesis, are transcribed later than most nodulation genes. As mutation of *fixF* leads to the abolition of rhamnose-rich polysaccharides and ineffective nodules on *V. unguiculata* (Jabbouri *et al*., 1996), the symbiotic role of this biosynthetic pathway could be examined using classical methods.

Experimental procedures

Molecular and microbiological techniques

E. coli recombinants were grown at 37°C in Luria–Bertani medium, Terrific broth or in two-fold YT medium (Sambrook *et al*., 1989). Strains of *Rhizobium* were raised at 27°C in/on *Rhizobium* minimal medium supplemented with succinate (RMM; Broughton *et al*., 1986). Flavonoid induction of *nod* genes was accomplished by adding daidzein to liquid cultures of NGR234 to a final concentration of 200 nM 1 h and 24 h before harvesting the cells. Cosmid or phage M13 DNA, as well as rhizobial genomic DNA, was prepared using standard procedures (Perret *et al*., 1991; Freiberg *et al*., 1997). Sequence analysis of M13 recombinants was performed as described previously (Freiberg *et al*., 1997).

Construction of the ORF map

The 536 165 bp sequence of pNGR234*a* was divided into 441 segments representing the 416 predicted genes as well as all intergenic regions. Segments were amplified using specific oligonucleotide pairs (primers of 18–20 bases) and cosmid (Perret *et al*., 1991) or M13 genomic DNA as templates in standard PCR reactions. Criteria for selecting the target sequences for PCR amplification included the following: (i) fragment size was limited to approximately 1 kb (smallest = 567 bp; to largest = 1953 bp; mean = 1121 bp) to minimize intensity differences in the subsequent hybridizations; (ii) adjacent primers were optimized to avoid overlapping PCR products and to minimize the distance between each amplified fragment (mean gap size = 95 bp); and (iii) for putative operons, 5′ primers were placed upstream of predicted transcription start points. In some cases, it was necessary to use inserts of M13 phages from the pSym sequence library were used in place of the PCR products. The 3′ extremities of M13 clones were verified by sequencing to avoid selection of co-ligation artefacts and to minimize overlap with adjacent, amplified fragments. Before Southern transfer, the size, quantity and overall quality of amplified products were verified. Standardized amounts of DNA were separated on 1% (w/v) agarose gels and vacuum blotted onto GeneScreen Plus nylon membranes (DuPont NEN). Separation of the target fragment on gels permitted accurate determination of the hybridizing band and elimination of false-positive results.

RNA extraction, labelling and hybridizations

Cultures of cells were collected by centrifugation when the absorbance at 600 nm reached 0.4–0.5. Bacteroids were isolated from effective nodules of *V. unguiculata* and *C. cajan* grown in large Leonard jars (Pueppke and Broughton, 1999) for 6–8 weeks after inoculation with NGR234. After collection, the nodules were crushed in liquid nitrogen and resuspended in sterile water. Debris was removed by filtration, and bacteroids were recovered by centrifugation ($4000 \times g$ for 5 min). Purification, labelling of RNA from bacteroids and rhizobia with [γ-^{32}P]-ATP as well as hybridization were performed as described previously (Fellay *et al*., 1995b), except that unlabelled competitor RNA was not added to the prehybridization solutions. Pools of DNA sequences purified from NGR234 by subtractive DNA hybridization against *R. fredii* USDA257 were also obtained and labelled with ^{32}P (Perret *et al*., 1994). Filters were exposed to X-ray films for 2–21 days depending on the intensity of the signal.

Acknowledgements

We wish to thank D. Gerber and S. Relić for their help in many aspects of this work. We are grateful to S. Jabbouri for critical reading of the manuscript. Financial assistance was provided by the Fonds National Suisse de la Recherche Scientifique (grant no. 31-45921.95) and the University of Geneva.

References

Berck, S., Perret, X., Quesada-Vincens, D., Promé, J.-C., Broughton, W.J., and Jabbouri, S. (1999) NolL of *Rhizobium* sp. strain NGR234 is required for *O*-acetyltransferase activity. *J Bacteriol* **181:** 957–964.

Broughton, W.J., Heycke, N., Meyer z.A., H., and Pankhurst, C.E. (1984) Plasmid linked *nif* and ''*nod*'' genes in fast-growing rhizobia that nodulate *Glycine max*, *Psophocarpus tetragonolobus*, and *Vigna unguiculata*. *Proc Natl Acad Sci USA* **81:** 3093–3097.

Broughton, W.J., Wong, C.H., Lewin, A., Samrey, U., Myint, H., Meyer z.A., H., *et al*. (1986) Identification of *Rhizobium* plasmid sequences involved in recognition of *Psophocarpus*, *Vigna*, and other legumes. *J Cell Biol* **102:** 1173–1182.

Cren, M., Kondorosi, A., and Kondorosi, E. (1995) NolR controls expression of the *Rhizobium meliloti* nodulation genes involved in the core Nod factor synthesis. *Mol Microbiol* **15:** 733–747.

D'Haeze, W., Gao, M.-S., De Rycke, R., Van Montagu, M., Engler, G., and Holsters, M. (1998). Roles for Azorhizobial Nod factors and surface polysaccharides in intercellular invasion and nodule penetration, respectively. *Mol Plant–Microbe Interact* **11:** 999–1008.

Dénarié, J., Debellé, F., and Promé, J.-C. (1996) *Rhizobium* lipochitooligosaccharide nodulation factors: signalling molecules mediating recognition and morpohogenesis. *Annu Rev Biochem* **65:** 503–535.

Ebeling, S., Kündig, C., and Hennecke, H. (1991) Discovery of a rhizobial RNA that is essential for symbiotic root nodule development. *J Bacteriol* **173:** 6373–6382.

Fellay, R., Rochepeau, P., Relić, B., and Broughton, W.J. (1995a) Signals to and emanating from *Rhizobium* largely

control symbiotic specificity. In *Pathogenesis and Host Specificity in Plant Diseases. Histopathological, Biochemical, Genetic and Molecular Bases*. Singh, U.S., Singh, R.P., and Kohmoto, K. (eds). Oxford: Pergamon Elsevier Science, pp. 199–220.

Fellay, R., Perret, X., Viprey, V., Broughton, W.J., and Brenner, S. (1995b) Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol Microbiol* **16:** 657–667.

Fellay, R., Hanin, M., Montorzi, G., Frey, J., Freiberg, C., Golinowski, W., *et al*. (1998) *nodD2* of *Rhizobium* sp. NGR234 is involved in the repression of the *nodABC* operon. *Mol Microbiol* **27:** 1039–1050.

Fischer, H.M. (1994) Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev* **58:** 352–386.

Fisher, R.F., and Long, S.R. (1993) Interactions of NodD at the *nod* box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. *J Mol Biol* **233:** 336–348.

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.

Girard, M., Flores, M., Brom, S., Romero, D., Palacios, R., and Dávila, G. (1991) Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. phaseoli. *J Bacteriol* **173:** 2411–2419.

Girard, L., Valderrama, B., Palacios, R., Romero, D., and Dávila, G. (1996) Transcriptional activity of the symbiotic plasmid of *Rhizobium etli* is affected by different environmental conditions. *Microbiology* **142:** 2847–2856.

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

Hanin, M., Jabbouri, S., Quesada-Vincens, D., Freiberg, C., Perret, X., Promé, J.-C., *et al*. (1997) Sulphation of *Rhizobium* sp. NGR234 Nod factors is dependent on *noeE*, a new host-specificity gene. *Mol Microbiol* **24:** 1119–1129.

Hanin, M., Jabbouri, S., Broughton, W.J., and Fellay, R. (1998) SyrM1 of *Rhizobium* sp. NGR234 activates transcription of symbiotic loci and controls the level of sulfated Nod factors. *Mol Plant–Microbe Interact* **11:** 343–350.

Hanin, M., Jabbouri, S., Fellay, R., Broughton, W.J., and Quesada-Vincens, D. (1999) Molecular aspects of host-specific nodulation. In *Plant–Microbe Interactions*. Stacey, G., and Keen, N.T. (eds). American Phytopathology Society, pp. 1–37.

Jabbouri, S., Hanin, M., Fellay, R., Quesada-Vincens, D., Reuhs, B., Carlson, R.W., *et al*. (1996) *Rhizobium* species NGR234 host-specificity of nodulation locus III contains *nod* and *fix* genes. In *Biology of Plant Microbe Interactions*. Stacey, G., Mullin, B., and Gresshoff, P.M. (eds). St Paul, MN: International Society for Molecular Plant–Microbe Interactions, pp. 319–324.

Kiss, E., Mergaert, P., Olàh, B., Kereszt, A., Staehelin, C., Davies, A.E., *et al*. (1998) Conservation of *nolR* in the *Sinorhizobium* and *Rhizobium* genera of the *Rhizobiaceae* family. *Mol Plant–Microbe Interact* **11:** 1186–1195.

Kondorosi, E., Gyuris, J., Schmidt, J., John, M., Duda, E., Hoffmann, B., *et al*. (1989) Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. *EMBO J* **8:** 1331–1340.

Lewin, A., Rosenberg, C., Meyer z.A., H., Wong, C.H., Nelson, L., Manen, J.-F., *et al*. (1987) Multiple host-specificity loci of the broad-host-range *Rhizobium* sp. NGR234 selected using the widely compatible legume *Vigna unguiculata*. *Plant Mol Biol* **8:** 447–459.

Machado, D., Pueppke, S.G., Vinardel, J.M., Ruiz-Sainz, J.E., and Krishnan, H.B. (1998) Expression of *nodD1* and *nodD2* in *Sinorhizobium fredii*, a nitrogen-fixing symbiont of soybean and other legumes. *Mol Plant–Microbe Interact* **11:** 375–382.

Perret, X., Broughton, W.J., and Brenner, S. (1991) Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234. *Proc Natl Acad Sci USA* **88:** 1923–1927.

Perret, X., Fellay, R., Bjourson, A.J., Cooper, J.E., Brenner, S., and Broughton, W.J. (1994) Subtraction hybridisation and shot-gun sequencing: a new approach to identify symbiotic loci. *Nucleic Acids Res* **22:** 1335–1341.

Perret, X., Viprey, V., Freiberg, C., and Broughton, W.J. (1997) Structure and evolution of NGRRS-1, a complex, repeated element in the genome of *Rhizobium* sp. strain NGR234. *J Bacteriol* **179:** 7488–7496.

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

Relić, B., Fellay, R., Lewin, A., Perret, X., Price, N.J.P., Rochepeau, P., *et al*. (1993a) *nod* genes and Nod factors of *Rhizobium* species NGR234. In *New Horizons in Nitrogen Fixation*. Palacios, R., and Mora, J. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 183–189.

Relić, B., Talmont, F., Kopcinska, J., Golinowski, Promé, J.-C., and Broughton, W.J. (1993b) Biological activity of *Rhizobium* sp. NGR234 Nod-factors on *Macroptilium atropurpureum*. *Mol Plant–Microbe Interact* **6:** 764–774.

Relić, B., Perret, X., Estrada-García, M.T., Kopcinska, J., Golinowski, W., Krishan, H.B., *et al.* (1994) Nod factors of *Rhizobium* are a key to the legume door. *Mol Microbiol* **13:** 171–178.

van Rhijn, P., and Vanderleyden, J. (1995) The *Rhizobium*–plant symbiosis. *Microbiol Rev* **59:** 124–142.

Rochepeau, P., Fellay, R., Jabbouri, S., Perret, X., and Broughton, W.J. (1997) Region II of *Rhizobium* sp. NGR234 inhibits nodulation of *Medicago sativa* by *R. meliloti nodIJ* and *nodQ1* mutants. *Mol Plant–Microbe Interact* **10:** 978–983.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

van Slooten, J.C., Cervantes, E., Broughton, W.J., Wong, C.H., and Stanley, J. (1990) Sequence and analysis of the *rpoN* sigma factor gene of *Rhizobium* sp. strain NGR234, a primary coregulator of symbiosis. *J Bacteriol* **172:** 5563–5574.

Trinick, M.J. (1980) Relationships amongst the fast-growing *Rhizobium* of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* sp., *Acacia farnesiana*, and *Sesbania grandiflora*

and their affinities with other *Rhizobium* groups. *J Appl Bacteriol* **49:** 39–53.

Vasse, J., de Billy, F., Camut, S., and Truchet, G. (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J Bacteriol* **172:** 4295–4306.

Viprey, V., Del Greco, A., Golinowski, W., Broughton, W.J., and Perret, X. (1998) Symbiotic implications of the type III protein secretion machinery in *Rhizobium*. *Mol Microbiol* **28:** 1381–1389.