Acta Biochimica Polonica

Vol. 48 No. 2/2001

359–365

QUARTERLY

Minireview

nod Genes and Nod signals and the evolution of the rhizobium legume symbiosis★

Frédéric Debellé[1✉], Lionel Moulin[2], Brigitte Mangin[3], Jean Dénarié[1] and Catherine Boivin[2]

[1]LBMRPM INRA-CNRS BP27 31326 Castanet Tolosan Cedex, France; [2]LSTM-IRD BP5035 34032 Montpellier Cedex, France; [3]BIA-INRA BP27 31326 Castanet Tolosan Cedex, France

Received: 14 February, 2001

Key words: rhizobium, legume, symbiosis, evolution, nodulation, Nod factors

The establishment of the nitrogen-fixing symbiosis between rhizobia and legumes requires an exchange of signals between the two partners. In response to flavonoids excreted by the host plant, rhizobia synthesize Nod factors (NFs) which elicit, at very low concentrations and in a specific manner, various symbiotic responses on the roots of the legume hosts. NFs from several rhizobial species have been characterized. They all are lipo-chitooligosaccharides, consisting of a backbone of generally four or five glucosamine residues N-acylated at the non-reducing end, and carrying various O-substituents. The N-acyl chain and the other substituents are important determinants of the rhizobial host specificity. A number of nodulation genes which specify the synthesis of NFs have been identified. All rhizobia, in spite of their diversity, possess conserved *nodABC* 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. The central role of NFs and *nod* genes in the Rhizobium-legume symbiosis suggests that these factors could be used as molecular markers to study the evolution of this symbiosis.

We have studied a number of NFs which are N-acylated by α,β-unsaturated fatty acids. We found that the ability to synthesize such NFs does not correlate with taxonomic position of the rhizobia. However, all rhizobia that produce NFs such nodulate plants belonging to related tribes of legumes, the Trifolieae, Vicieae, and Galegeae, all of them being members of the so-called galegoid group. This suggests that the ability to recognize the NFs with α,β-unsaturated fatty acids is limited to this group of legumes, and thus might have appeared only once in the course of legume evolution, in the galegoid phylum.

★Presented at the International Conference on "Molecular Architecture of Evolution, Primary and Secondary Determinants" Poznań, Poland, October 29–31, 2000.

✉Corresponding author: tel: (33) 5612 85463, fax: (33) 5612 85061, e-mail: debelle@toulouse.inra.fr

Abbreviation: NF, Nod factor.

Nod factors (NFs) are difficult to characterize biochemically, their structure can be determined only for a limited number of rhizobial strains. We therefore wanted to assess whether *nod* gene sequence, which is easier to obtain, could give clues on the NF structure. We focused on the *nodA* gene, which is present in a single copy in all rhizobia, and whose product, an NF acyl transferase, interacts with two substrates, an acyl chain donor and a substituted chito oligomeric acceptor. These two substrates vary in structure among rhizobia, and might influence the NodA structure and sequence. We sequenced therefore the entire *nodA* gene of 36 strains whose NF structure have been characterized. Phylogenetic analysis of the NodA sequences showed that they form clusters which do not correlate with rhizobial taxonomic position. Instead, a correlation could be found between NodA sequence and structural features of the NF such as O-fucosylation, O-arabinosylation, or N-acylation by α,β-unsaturated fatty acids. Four structural types of NF were distinguished, based on the structure of the *N*-acyl chain and the type of O-glycosylation. The correlation between NodA sequence and NF structural type was confirmed by a statistical analysis which identified amino-acid residues informative on the NF type and provided a tool to predict the NF type on the basis of the *nodA* gene sequence. This tool will be useful to look for novel NF structures and to study the evolution of NF structure in the course of legume evolution.

The symbiotic relationship between rhizobium bacteria and legumes results in the formation on the roots of the host plant of differentiated organs called nodules in which the bacteria reduce atmospheric nitrogen into ammonia. Ammonia is used by the host plant which in exchange provides the rhizobia with carbon sources.

The rhizobium-legume symbioses are highly specific, each rhizobium infecting and nodulating defined legume plants. Nevertheless, the degree of specificity is variable. Some rhizobia such as *Sinorhizobium meliloti* or *Rhizobium leguminosarum* bv *trifolii* are specific for a few legume genera, the former for *Medicago*, *Melilotus* and *Trigonella* the latter for *Trifolium*. Other bacteria, like *R.* sp NGR234 can infect legumes in more than 120 genera and even a non-legume, *Parasponia*. Although both rhizobia and legumes can survive in the absence of a symbiotic partner, the tight association between plant and rhizobium in the symbiotic stage, and the specificity of the interactions, suggest that the host and symbiont could have evolved in parallel.

POLYPHYLETIC ORIGIN OF RHIZOBIA

Comparison of host plant and rhizobial phylogenies does not support their cospeciation. Rhizobium phylogeny, based mainly on the analysis of 16S rDNA sequences, indicates that rhizobia belong to the four major lineages of α-proteobacteria (Fig. 1), which contain also non symbiotic bacteria. Thus some rhizobia can be more closely related to non symbionts than to other rhizobia, suggesting a polyphyletic origin of these bacteria. Rhizobia belonging to different phylogenetic branches are able to nodulate the same legume species. For example distantly related *Bradyrhizobium japonicum*, *B. elkanii* and *S. fredii* nodulate soybean. Similarly, *Azorhizobium caulinodans*, *S. terangae* and *S. saheli* nodulate *Sesbania*. In addition, strains of the *S. terangae* or *S. saheli* species nodulate the very distant *Acacia* (Mimosoideae subfamily) and *Sesbania* (Papilionoideae subfamily) legumes. This indicates that there is little correlation between rhizobium and host plant phylogenies (Doyle, 1998), and that analysis of 16S rDNA or housekeeping gene sequences (Turner & Young, 2000) are unlikely to provide evidence of coevolution of the symbiotic partners. We therefore investigated genes

that are directly involved in nodulation and determinationof host specificity.

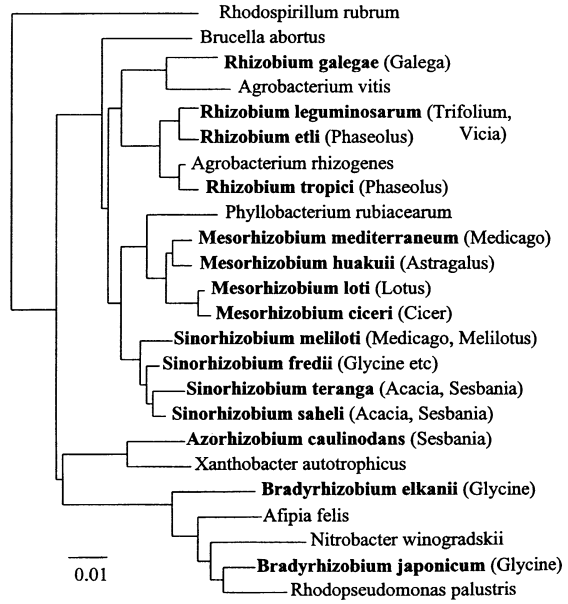

Fig ure 1. 16S rDNA phylo gen etic tree of rhizobia (bold let ters) and re lated bac te ria.

THE NODULATION GENES HAVE A MONOPHYLETIC ORIGIN AND SPECIFY Nod FACTOR BIOSYNTHESIS

All rhizobia identified so far carry nodulation (*nod*, *noe*, *nol*) genes which are re quired for infection and nodule organogenesis (Downie, 1998). Some of these genes, such as *nodABCD* are found in all rhizobia while other are present only in some species. Sequence anal y sis of the *nod* genes has shown that they are highly conserved even between distantly related lineages of rhizobia, suggesting that they might have a monophyletic origin and could have been transmitted to different groups of non-symbiotic bacteria by horizontal trans fer. For ex am ple, the *nodD* genes belong to the *lysR* fam ily of transcriptional ac tivators and the *nodD* genes from all rhizobia are more closely related to each other than they are to any other mem ber of the *lysR* family. The nodulation genes are involved in an exchange of signals between legume host plants and rhizobia (Dénarié *et al*., 1996; Perret *et al.*, 2000). Legume plants secrete into the rhizosphere secondary metabolites, mainly flavonoids, which are thought to in teract with the rhizobial regulatory NodD protein to activate the expression of the rhizobium nodulation genes. These genes specify the biosynthesis and secretion of lipo-chitooligosaccharidic molecules, the Nod factors, which at very low concentrations can induce in host plants symbiotic responses such as root hair deformation, cortical cell division and nodule primordium formation. Although all NFs are lipo-chitooligosaccharides, NFs from dif fer ent rhizobia dif fer in the sub stituents of the chitooligosaccharide backbone, which confer to them specificity towards a sub set of le gume plants. The *nodABCD* genes specify the synthesis of the lipooligosaccharide core of all NFs and are strictly required for nodulation. Host specificity genes, such as *nodL* or *nodH,* spec ify the var i ous NF substitutions. Mutations in these genes can re sult in changes in the rhizobial host range. Therefore *nod* genes and NFs play a central role in nodulation and host range de ter mi nation. Thus, we won dered whether they could be used as molecular markers to follow the evolution of the rhizobium-legume symbioses. In particular, we wanted to assess whether NF struc tural fea tures could be as sociated with groups of phylogenetically related rhizobia and/or legume host plants. This might give clues on how the mechanisms of rec og ni tion be tween sym bi otic partners evolved from ancient to modern legumes. In addition, this could shed light on the molecular mechanisms of coevolution of the symbioticpartners.

LEGUMES NODULATED BY RHIZOBIA PRODUCING NFs WITH α,β-UNSATURATED FATTY ACIDS ARE PHYLOGENETICALLY RELATED

Most rhizobia pro duce NFs acylated by fatty acids from the general lipid metabolism,

which are saturated fatty acids (C18:0: stearic acid, C16:0: palmitic acid) or fatty acids carrying one *cis* double bond (C18:1: vaccenic acid). However, some rhizobia secrete NFs that contain α,β-unsaturated fatty acids i.e. fatty acids carrying *trans* double bonds conjugated to the carbonyl group. When present, such fatty acids are important determinants of the host range. Their biosynthesis requires the host specific *nodFE* genes and their transfer to the NF chitooligosaccharide backbone depends on particular *nodA* genes (Debellé *et al*., 1996; Ritsema *et al*., 1996). Such α,β-unsaturated fatty acids were previously known to substitute the NF of *S. meliloti*, *R. leguminosarum* bv *viciae* and *R. l.* bv *trifolii.* More recently we have identified them on the NF of *R. galegae*, *Mesorhizobium huakuii* (Yang *et al.,* 1999) and *M.* sp N33 (*Oxytropis arctobia*) (Poinsot *et al.,* unpublished). The α,β-unsaturated fatty acids produced by these various strains differ in their chain length, number of conjugated double bonds and substitutions (Fig. 2). Biosynthesis of these fatty acids is achieved by rhizobia belonging to various taxonomic groups: *Sinorhizobium*, *Rhizobium*, *Mesorhizobium* which include also strains synthesizing NF with general metabolism fatty acids. Therefore the ability to synthesize α,β-unsaturated fatty acids is not characteristic of a particular taxonomic group of rhizobia. On the contrary, the legume hosts of the rhizobia producing α,β-unsaturated fatty acids all belong to a group of phylogenetically related tribes, the so-called galegoid group, which includes the Trifolieae, Vicieae, and Galegeae tribes (Fig. 3). Our interpretation of these results is that most extant legumes, like archaic legumes, do not recognize NFs with α,β-unsaturated fatty acids but interact with NFs substituted by general metabolism fatty acids. The ability to recognize α,β-unsaturated fatty acids appeared once in legume evolution, in an ancestor of the galegoid plants. In parallel, rhizobia of several taxonomic groups (excluding *Azorhizobium* and *Bradyrhizobium*) acquired by horizontal transfer *nodFE* and *nodA* genes allowing the synthesis of NFs with α,β-unsaturated fatty acids. Then there occurred a diversification of the NF recognition mechanism allowing the plants to distinguish α,β-unsaturated fatty acids differing in chain length, number of unsaturations or substitutions. Simultaneously, allelic diversification of *nodFE* and *nodA* in rhizobia allowed the synthesis of the various α,β-unsaturated fatty acids.

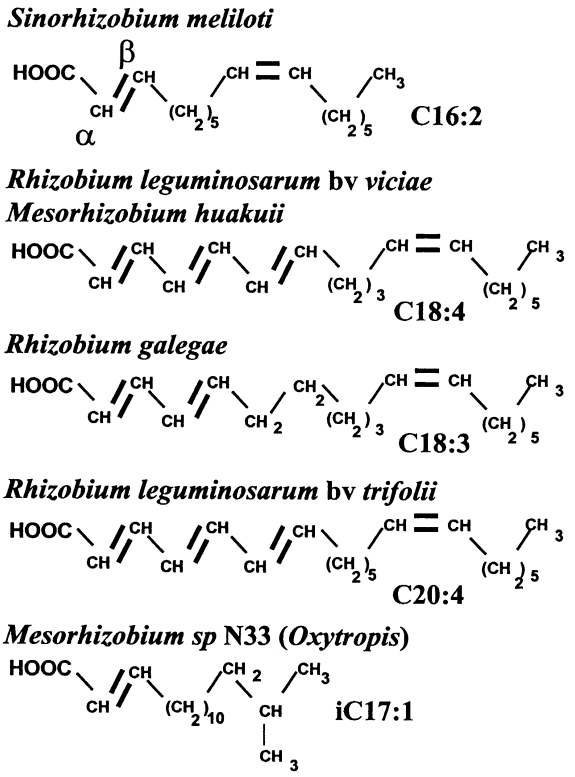

Figure 2. *nodAFE*-dependent α,β-unsaturated acyl substituents of Nod factors produced by various rhizobia.

OTHER CORRELATIONS BETWEEN LEGUME PHYLOGENY AND TYPE OF NF PRODUCED BY SYMBIOTIC BACTERIA

There are other examples of an original Nod factor substitution associated with a specific phylogenetic group of host legumes. One is

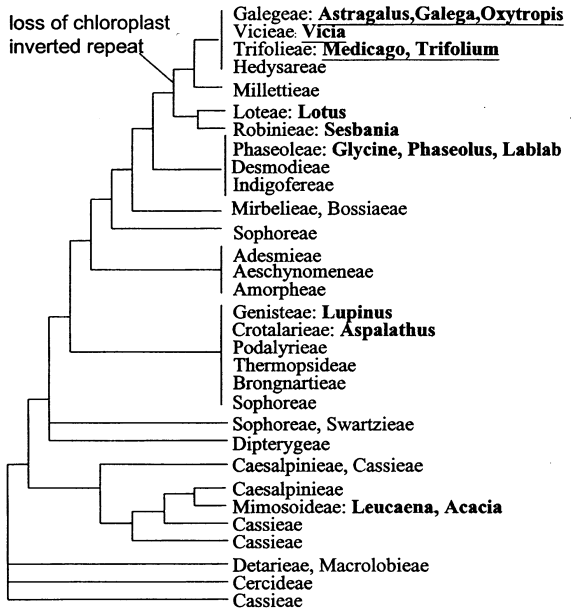

Figure 3. Phylogenetic tree of legume based on the analysis of rbcL sequence (after Doyle, 1998).

A cluster of phylogenetically related legumes (underlined) is nodulated by rhizobia producing Nod factors substituted by α,β-unsaturated fatty acids.

that of the arabinose found on NFs of *Sesbania* nodulating rhizobia. While many legumes are nodulated by rhizobia producing NFs 6-O substituted by a fucosyl group at the reducing end, only rhizobia nodulating *Sesbania* sp. in the Robinieae tribe produce NFs with a 3-*O* arabinosyl group at the reducing end, in addition to the fucosyl group (Lorquin *et al.,* 1997). These rhizobia – *Azorhizobium caulinodans*, *S. saheli* bv *sesbaniae* and *S. terangae* bv *sesbaniae* strains – belong to taxonomically different groups but all of them produce the same NFs which so far have not been isolated from rhizobia nodulating other legumes. It is therefore tempting to hypothesize that, in the course of evolution, *Sesbania* plants have acquired the unique ability to recognize the arabinose substituents on NFs and select rhizobia on this basis. The various *Sesbania*-nodulating rhizobia are likely to have acquired the genes required for the NF arabinosyl substitution by horizontal transfer.

Some substitutions are found on NFs of rhizobia which nodulate plants belonging to distinct phyla of legumes. One example is that of 6-*O* sulfation of the reducing end in *S. meliloti, M. huakuii*, *M.* sp N33, and *R. tropici* NFs. This substitution is an important determinant of the host range (Roche *et al*., 1991). Rhizobia producing 6-*O* sulfated NFs nodulate plants of different tribes (Trifolieae, Galegeae, Phaseoleae, Acacieae) which also comprise legumes nodulated by rhizobia producing non-sulfated NFs. Thus it is likely that the ability to recognize the *O*-sulfate substituent was acquired (or lost) several times in the course of legume evolution.

Finding of a correlation between legume phylogeny and type of the NFs produced by symbiotic bacteria is more difficult in the case when the symbiotic associations are less specific. For example, *Phaseolus* species can be nodulated by a variety of rhizobial strains which produce at least two types of structurally different NFs: *R. etli* produces NFs carrying a fucosyl group at the reducing end whereas *R. tropici* synthesizes NFs 6-*O* sulfated at the reducing end. One explanation for these observations is that *Phaseolus* carries non stringent receptors able to recognize both types of NFs. Alternatively, these legumes could carry two types of receptors.

Some rhizobia belonging to different cross inoculation groups have been shown to produce similar Nod factors. Thus, the determinants of the host range other than NF/receptor interaction have to be taken into account. For example, specific interactions between NodD and flavonoids, and type III secretion systems have been shown to play a role in the host range determination (Perret *et al.,* 2000).

A major problem in trying to follow the evolution of the mechanisms of recognition between symbiotic partners is the lack of NF structural data for many symbiotic relationships. NF structure determination requires heavy work, and so far NF structures are known for symbionts of only about a hundred among the 16 000 species of legumes. Since

molecular analysis of DNA is much easier to carry out on a large scale, we attempted to use *nod* gene sequence analysis to gain insight into NF structures.

nod GENE SEQUENCE ANALYSIS AS A TOOL TO PREDICT NF STRUCTURE

The genes involved in NF biosynthesis can be classified in two groups: the *nodABC* genes are present in all rhizobia and are responsible for the synthesis of the lipooligosaccharide core common to all Nod factors; the host specificity genes are responsible for the biosynthesis of the various substituents and their transfer to the oligosaccharide backbone of NFs. Thus characterization of the host specificity genes seems to be the most direct way to have access to NF structure. However not all of the genes responsible for the various substitutions are known. In addition, due to frequent rearrangements in rhizobial genomes, truncated *nod* genes and *nod* promoters have been observed (Krishnan *et al.,* 1992). Inactive *nod* genes would make prediction of NF structure based on the presence of specific *nod* genes unreliable. We thus focused on the *nodABC* genes which are found in a single copy in all rhizobial strains. Among them, *nodA* which specifies the transfer of an acyl chain to the oligosaccharide backbone of NFs, appeared the most likely to provide information on the structure of NFs. The NodA protein interacts with two substrates, an acyl chain donor and an acyl chain acceptor which is a substituted chitin oligosaccharide. Previous work had shown that different NodA proteins exhibit different specificity for the acyl chains (Debellé *et al.,* 1996; Ritsema *et al.,* 1996). These differences in specificity might be reflected in differences in NodA structure and, thus, in the amino-acid sequence of NodA. Similarly putative differences in specificity toward various acyl chain acceptors might be reflected in NodA amino-acid sequences. Substituents of the NF oligosaccharide backbone that are added before NF acylation, such as fucose, are most likely to be recognized by NodA.

We therefore determined the *nodA* sequences, when they were not available in data bases, of all rhizobia whose NF structure is known, and which belong to all known taxonomic branches of rhizobia. Similarities between the NodA proteins were investigated by a phylogenetic analysis and we attempted to correlate NodA phylogenetic clusters with structural features of NFs. These correlations were further validated by a statistical method.

NodA SEQUENCE AS A TOOL TO PREDICT NF TYPE AND TO STUDY RHIZOBIUM-LEGUME COEVOLUTION

Phylogenetic analysis grouped NodA sequences in eight clusters. One cluster included all *Bradyrhizobium* and a second all *Azorhizobium* strains, suggesting that this clustering reflects taxonomic distance between strains. The NodA of other rhizobia spread into six clusters, half of them grouping different genera. We were able to associate each cluster with NF substitutions shared by all members of the group. This allowed us to define four major NF structural types: the F type corresponding to NFs substituted by a fucose derivative at the reducing end, the A type corresponding to NFs arabinosylated and fucosylated at the reducing end, the U type corresponding to NFs substituted by α,β unsaturated fatty acids at the non-reducing end, the S type with none of the three above mentioned substitutions.

We attempted to validate the grouping of the NodA sequences in four groups corresponding to the four NF types by using a statistical method. For this purpose we first identified informative positions in the NodA sequence, i.e. positions for which a good correlation between the nature of the amino-acid residue and the NF type was obtained (Moulin *et al.,*

unpublished). Using amino-acid counts at these positions the probability of the four NF types could then be computed, allowing the prediction of the NF type for a given NodA sequence (Moulin *et al.,* unpublished). The method was validated by cross validation: in most cases, the predicted NF type corresponded to what was known from biochemical analysis.

We used the above described method together with phylogenetic analysis to predict the NF type for rhizobial strains recently described for which no biochemical analysis of NFs had been performed. In most cases the prediction was in good agreement with what we expected knowing the host range of the strain. NodA sequence analysis could thus be used, instead of 16S rDNA analysis, to characterize the symbiotic phenotype of newly described rhizobium strains isolated from various ecosystems.

Analysis of NodA sequence together with additional NF structure determination should also allow us to follow the evolution of NF structures in the course of the legume-rhizobium coevolution, from primitive to more modern symbiotic associations.

REFERENCES

Debellé, F., Plazanet, C., Roche, P., Pujol, C., Savagnac, A., Rosenberg, C., Promé, J.C. & Dénarié, J. (1996) The NodA proteins of *Rhizobium meliloti* and Rhizobium tropici specify the N-acylation of Nod factors by different fatty acids. *Mol. Microbiol*. **22**, 303–314.

Dénarié, J., Debellé, F. & Promé, J.C. (1996) Rhizobium lipo-chitooligosaccharide nodulation factors: Signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem*. **65**, 503–535.

Downie, J.A. (1998) Functions of rhizobial nodulation genes; in *The Rhizobiaceae* (Spaink, H.P., Kondorosi, A. & Hooykaas, P.J.J., eds.) pp. 387–402.

Doyle, J.J. (1998) Phylogenetic perspectives on nodulation: Evolving views of plants and symbiotic bacteria. *Trends Plant Sci*. **3,** 473–478.

Krishnan, H.B., Lewin, A., Fellay, R., Broughton, W.J. & Pueppke, S.G. (1992) Differential expression of *nodS* accounts for the varied abilities of *Rhizobium fredii* USDA257 and *Rhizobium* sp. strain NGR234 to nodulate *Leucaena* spp. *Mol. Microbiol*. **6**, 3321–3330.

Lorquin, J., Lortet, G., Ferro, M., Méar, N., Dreyfus, B., Promé, J.C. & Boivin, C. (1997) Nod factors from *Sinorhizobium saheli* and *S. teranga* bv. *sesbaniae* are both arabinosylated and fucosylated, a structural feature specific to *Sesbania rostrata* symbionts. *Mol. Plant-Microbe Interact.* **10**, 879–890.

Perret, X., Staehelin, C. & Broughton, W.J. (2000) Molecular basis of symbiotic promiscuity. *Microbiol. Mol. Biol. Rev*. **64**, 180–201.

Ritsema, T., Wijfjes, A.H.M., Lugtenberg, B.J.J. & Spaink, H.P. (1996) *Rhizobium* nodulation protein NodA is a host-specific determinant of the transfer of fatty acids in Nod factor biosynthesis. *Mol. Gen. Genet.* **251,** 44–51.

Roche, P., Debellé, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Dénarié, J. & Promé, J.-C. (1991) Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. *Cell* **67**, 1131–1143.

Turner S.L. & Young, J.P. (2000) The glutamine synthetases of rhizobia: Phylogenetics and evolutionary implications. *Mol. Biol. Evol.* **17**, 309–319.

Yang, G.-P., Debellé, F., Savagnac, A., Ferro, M., Schiltz, O., Maillet, F., Promé, D., Treilhou, M., Vialas, C., Lindstrom, K., Dénarié, J. & Promé, J.C. **(**1999) Structure of the *Mesorhizobium huakuii* and *Rhizobium galegae* Nod factors: A cluster of phylogenetically related legumes are nodulated by rhizobia producing α,β-unsaturated fatty acids. *Mol. Microbiol.* **34**, 227–237.