JOURNAL OF BACTERIOLOGY, Jan. 1995, p. 468–472
0021-9193/95/$04.00+0

Vol. 177, No. 2

# Phylogeny of Sym Plasmids of Rhizobia by PCR-Based Sequencing of a *nodC* Segment

TOHRU UEDA,* YUKO SUGA, NOBUTAKA YAHIRO, AND TATSUHIKO MATSUGUCHI

*Laboratory of Soil Microbiology and Biochemistry, Department of Agricultural Chemistry, Kyushu University, Fukuoka, Japan*

Received 12 July 1994/Accepted 3 November 1994

**To understand the host specificity of rhizobia and the relationship between the evolution of Sym plasmids and that of host plants, we determined partial *nodC* sequences of 10 representative rhizobium strains and then constructed an evolutionary tree for the deduced amino acid sequences with four published sequences. These coding sequences yield a phylogenetic tree similar to that for leghemoglobin of host plants, suggesting that the evolution of common nodulation genes may be linked to host legume evolution and speciation.**

Plants coexist with a great number of soil bacteria. Rhizobium-legume interaction stands out from other plant-microbe interactions as one in which a true developmental mutualism occurs. The ability of rhizobial strains to form effective nodules in which they reduce atmospheric nitrogen to ammonia and supply the plant with nitrogenous compounds is limited to certain host plants (2, 5, 27). For example, *Vicia* and *Pisum* spp. are host plants for *Rhizobium leguminosarum* bv. viciae, *Phaseolus* spp. are hosts for *R. leguminosarum* bv. phaseoli, *Trifolium* spp. are hosts for *R. leguminosarum* bv. trifolii, *Medicago* spp. are hosts for *Rhizobium meliloti*, *Glycine* spp. are hosts for *Bradyrhizobium japonicum*, and the tropical legume *Sesbania rostrata* is the host for *Azorhizobium caulinodans*. The nodulation events occur when the rhizobia respond to the presence of specific plant flavonoids that stimulate the coordinate expression of bacterial nodulation genes (*nod* genes). The *nod* genes, in turn, encode enzymes involved in the synthesis of Nod factors which act as determinants for host specificity and lead to formation of nitrogen-fixing root nodules (2, 5). The *nodABC* genes have been characterized as common *nod* genes, which are essential for nodulation to occur. Since NodC is homologous to chitin synthase (1) and acts as an *N*-acylglucosaminyltransferase (7), this protein is involved in the formation of the Nod factor backbone.

Recently, Young and Johnston reported that the phylogenetic tree for NodD protein does not correspond with the 16S rRNA phylogeny and suggested lateral gene transfer of Sym plasmids (25, 27). Oyaizu et al. also suggested lateral gene transfer, since the phylogeny of 16S rRNA of rhizobia does not relate to the host specificity of rhizobia (16).

These reports prompted us to study the relationship between the evolution of Sym plasmids and that of host plants by analysis of the *nodC* sequences. Since some rhizobia have some copies of *nodD*, analysis of *nodD* sequence may not be appropriate for the phylogenetic study of the plasmids. On the other hand, because all rhizobia contain only a single copy of *nodC* and this molecule has a role in the growth of the backbone of Nod factor (1, 7), we determined the *nodC* sequences of 10 representative rhizobium strains for the phylogenetic study of the plasmids. Our resulting phylogenetic trees of NodC protein suggest that the Sym plasmids evolved concordantly with host plant species divergence on an evolutionary timescale. In this report, we propose the hypothesis that coevolution has occurred not in rhizobium-legume interaction, but in Sym plasmid-legume interaction.

**Bacterial strains.** Bacterial strains used in this study are listed in Table 1. The rhizobium strains were grown in yeast mannitol medium. All strains were grown at 30°C with shaking.

**Cloning and sequencing of *nodC* segments.** After cultivation, about 100 mg of harvested cells was resuspended in 50 mM Tris–20 mM EDTA (final concentration, 1.5%) and lysed by incubation at 65°C for 15 min. The lysate was treated with RNase A and proteinase K followed by chloroform extraction and isopropanol precipitation. Crude DNA was purified by phenol extraction, chloroform extraction, and isopropanol precipitation. The primers for PCR amplification (17) were chosen by careful inspection of the four published rhizobial *nodC* sequences available in the GenBank, EMBL, and DDBJ DNA databases. The region amplified is a relatively variable part of the molecule and thus has a high density of information. The amino acid sequence deduced for this region is proposed to be an extracellular domain (9). The sequences of the amplification primers are as follows: 251F, AYGTIGTYGAYGAYGG WTC; 566R, AGCCARTACTCCATGTCGATCAA. The *nodC* segments were amplified from bacterial DNA by PCR. Nega-

TABLE 1. Bacterial strains used in this study

| Species and strain[a] | DDBJ accession no. |
|---|---|
| *Bradyrhizobium japonicum* USDA 6 | D28962 |
| *Bradyrhizobium elkanii* USDA 46 | D28963 |
| *Bradyrhizobium elkanii* USDA 61 | D28964 |
| *Bradyrhizobium elkanii* USDA 94 | D28965 |
| *Bradyrhizobium japonicum* USDA 122 | D28956 |
| *Bradyrhizobium japonicum* USDA 136 | D28957 |
| *Bradyrhizobium japonicum* USDA 142 | D28958 |
| *Rhizobium leguminosarum* bv. trifolii USDA 2161 | D28959 |
| *Rhizobium leguminosarum* bv. viciae USDA 2478 | D28960 |
| *Rhizobium leguminosarum* bv. phaseoli USDA 2676 | D28961 |

[a] USDA, U.S. Department of Agriculture, Beltsville, Md.

* Corresponding author. Mailing address: Laboratory of Soil Microbiology and Biochemistry, Department of Agricultural Chemistry, Kyushu University, Hakozaki 6-10-1, Fukuoka 812, Japan. Phone: 092-641-1101 ext. 6421. Fax: 092-632-1960.

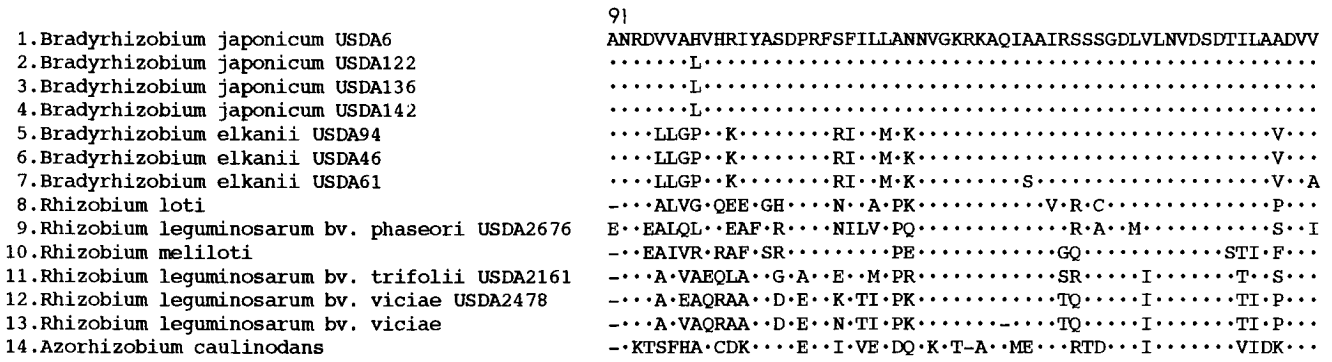
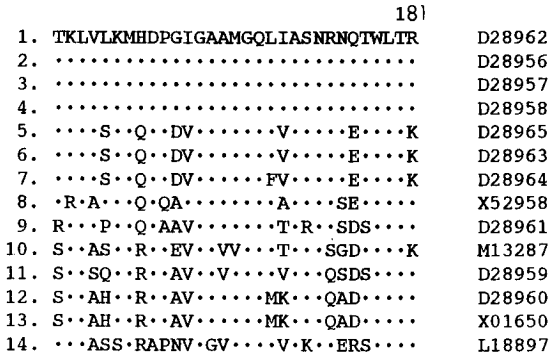



FIG. 1. Amino acid alignment of NodC proteins. Dots indicate identity, and hyphens indicate gaps. The numbering at the top is that for *R. meliloti* NodC. Database accession numbers are also indicated.




tive controls (water instead of DNA) showed no amplification. The reaction conditions were as follows: 10 ng of template DNA; 0.5 min at 37°C, 2 min at 60°C, and 0.5 min at 94°C for the first 5 cycles; and 0.5 min at 50°C, 0.5 min at 72°C, and 0.5 min at 94°C for the following 27 cycles. Products were purified, and the *nodC* fragments were then cloned into pT7BlueT vector by using *Escherichia coli* JM109. Analysis of double-stranded DNA with the universal M13 primer and the T7 primer was done by using the *Taq* Dye Primer Cycle Sequencing Kit and the DNA Sequencer model 373A (Applied Biosystems, Foster City, Calif.) in accordance with the manufacturer's directions. All of the nucleotides were confirmed by sequencing the complementary chain of the DNA. We independently cloned and sequenced three PCR fragments from each strain to make sure that they were the same.

**Phylogenetic analysis of NodC protein.** The four published *nodC* sequences are now available in the GenBank, EMBL, and DDBJ DNA databases, and we have determined partial sequences (of a 274-base segment amplified by PCR primers 251F and 566R) for 10 representative rhizobium strains by PCR (17). The deduced amino acid sequences were then compared with known sequences. As a result, none of the 10 analyzed sequences is identical to a published sequence. When the 14 sequences were aligned, several regions were found to be conserved in all sequences considered, whereas others were found to be divergent, as shown in Fig. 1.

Figure 2A shows an evolutionary tree constructed from such data analyzed by the neighbor-joining tree construction (NJ) method (18). The NJ method, which does not require the assumption of a constant substitution rate, appears to be one of the most effective tree construction methods available at the moment (14). In this analysis, the root of the tree was taken as the midpoint of the longest path, which was between *A. caulinodans* and a cluster of other rhizobia. To examine whether this tree topology remains the same since the rate of nucleotide substitution has not changed during evolution, we also constructed a phylogenetic tree by the unweighted pair group method using arithmetic averages (UPGMA) (13, 21) as shown in Fig. 2B. Interestingly, this tree has almost the same topology as the tree constructed by the NJ method. The reliability of the tree nodes was analyzed by using the bootstrapping program (4, 11). The percentages of 10,000 bootstrap resamplings that support each topological element are indicated in Fig. 2. Furthermore, we also constructed the NJ and UPGMA trees by use of the numbers of nonsynonymous substitutions (15) using the MEGA package (data not shown), and this showed that both trees have the same topology as the tree constructed by the NJ method with amino acid sequence data (Fig. 2A). These data suggest that the topology shown in Fig. 2A might be reliable and that *nodC* genes evolved at a relatively constant rate.

**Phylogenetic analysis of NodA protein and 16S rRNA.** We have also constructed phylogenetic trees by the NJ method from NodA protein sequence data using the public databases, as shown in Fig. 3. Though data for only five published sequences are available at the moment, we found that the topology of an evolutionary tree of NodA protein is quite similar to that of NodC protein. Furthermore, this NodA tree also has the same topology as the NodD tree constructed by other workers (19, 27). This analysis suggests that *nodA*, *nodC*, and *nodD* genes have evolved in a similar manner, which agrees with the fact that the *nodABCD* genes are linked in all known rhizobia. On the other hand, the topology of an evolutionary tree of 16S rRNA is quite dissimilar to that of NodC protein, as shown in Fig. 4. These data apparently suggest the lateral gene transfer of Sym plasmids, as reported by other workers (16, 25, 27).

**Phylogenies of NodC protein and leghemoglobin in host plants.** In order to understand the host specificity of rhizobia, we compared the topology of the NodC tree in Fig. 2A with that of leghemoglobin in host plants, as shown in Fig. 5. Our leghemoglobin phylogeny presented in Fig. 5 is consistent with the current hypothesis of legume phylogeny based on some sequences of nuclear genes (3). Interestingly, the topology of the evolutionary tree of NodC protein is similar to that of leghemoglobin in host plants. These observations, together with the fact that strains of *Rhizobium* species have the nodulation genes on the Sym plasmids (2, 5), suggest that the evolution of common nodulation genes may be linked to legume

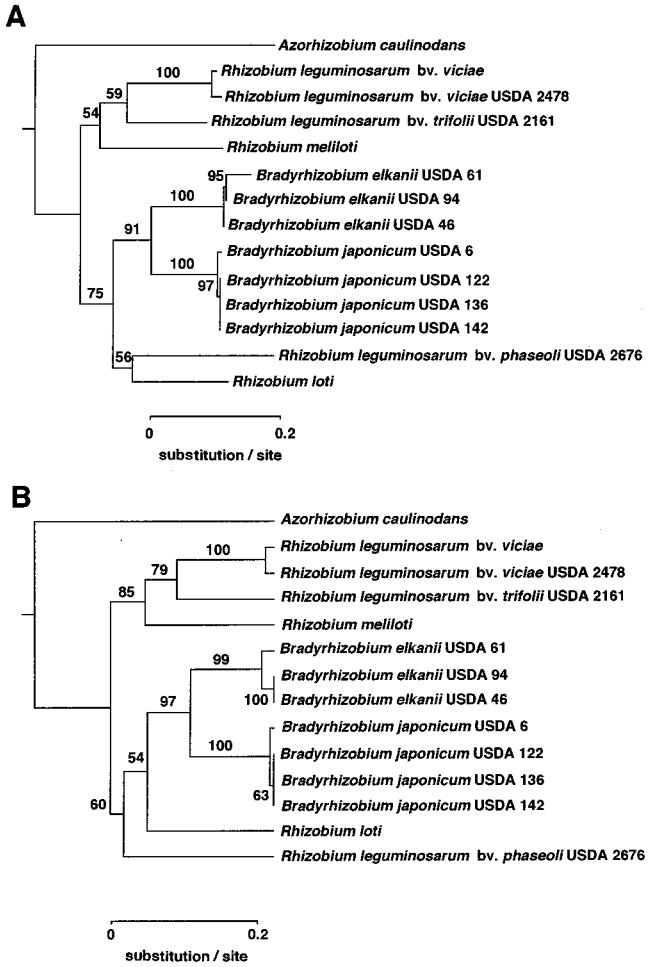



FIG. 2. Phylogeny of NodC protein sequences by the NJ method (A) and UPGMA (B). All four known different *nodC* sequences, corresponding to positions 270 to 543 in the *R. meliloti* sequence, were extracted from the GenBank, EMBL, and DDBJ DNA databases by using the ODEN system (8). The partial *nodC* gene sequences were then translated and aligned with each other to maximize the sequence similarity by using the ODEN multiple sequence alignment program (8) and by visual inspection. In estimating evolutionary distances, we discarded the regions common to the 14 sequences where deletions are observed even for one sequence, since there is no way to assess the rate and size of deletions during the course of evolution. Omission of these regions left us 87 comparable amino acid sites. To estimate evolutionary distances between amino acid sequences, we used Poisson correction. By using these estimates, molecular evolutionary trees were constructed from the resulting distance matrix by the NJ method (18) and UPGMA (13, 21) with the MEGA package (8). The percentages of 10,000 bootstrap resamplings that support each topological element are indicated (4).

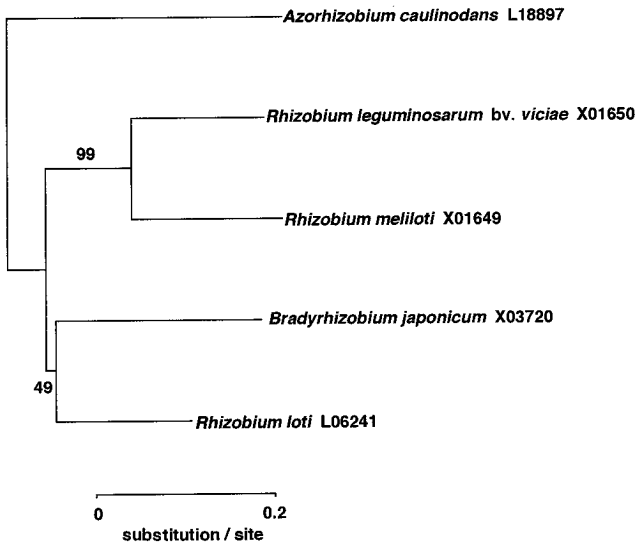



FIG. 3. Phylogeny of NodA protein sequences by the NJ method. All five known *nodA* sequences were extracted from the GenBank, EMBL, and DDBJ databases by using the ODEN system (8). The sequences were translated and then aligned with each other by visual inspection to maximize the sequence similarity. In estimating evolutionary distances, we discarded regions common to the five sequences where deletions are observed even for one sequence. Omission of the regions left us 151 comparable amino acid sites. To estimate evolutionary distances between nucleotide sequences, we used Poisson correction. By using these estimates, molecular evolutionary trees were constructed from the resulting distance matrix by the NJ method (18) with the MEGA package (11). The database accession numbers are indicated after the bacterial names. The percentages of 10,000 bootstrap resamplings that support each topological element are indicated (4).

evolution and speciation. In other words, mutual adaptations of Sym plasmids and legumes have become refined in the course of their association; thus, common nodulation genes and their host plants appear to radiate in parallel, or cospeciate. This pairwise coevolution may be related to interaction of Nod factors and corresponding receptor proteins in legumes on an evolutionary timescale (12). It is reported that the study of coevolution must be the analysis of reciprocal genetic changes that might be expected to occur in two or more ecologically interacting species, and very strong evidence for coevolution of two groups can be obtained if the phylogenetic trees of the two groups are congruent or nearly so (6). Recently it has also been suggested that the evolution of small-DNA-virus families may be linked to vertebrate host evolution and speciation (20). To our knowledge, our experiments provide the first molecular evidence that coevolution might have occurred in plant-plasmid interaction. However, it must be noted that some rhizobium strains, especially within the genus *Bradyrhizobium*, have broad host ranges (2, 5). On the other hand, strains within the genus *Rhizobium* tend to have rather narrow host ranges, so we might reasonably look for signs of parallel evolution of Sym plasmids in these strains and their host plants. Furthermore, it must also be noted that the *Bradyrhizobium* and *Azorhizobium* genes are all on the chromosome. There might be some possibility that nodulation genes on Sym plasmids had been integrated into their chromosome on an evolutionary timescale.

Assuming that coevolution has occurred in plasmid-legume interaction, Fig. 5 suggests that the rate of NodC evolution is similar (within an order of magnitude) to that of leghemoglobin evolution in host legumes. It is worthwhile to note that the first ancestors of legumes are considered to have appeared in the late Cretaceous-early Tertiary period, 60 to 70 million years before the present (22). Although this would be an interesting opportunity to apply a plant timescale to bacterial plasmid evolution, we cannot calculate the rate of leghemoglobin and NodC protein evolution because we have no available data at the moment about divergence time between the tribe Genisteae (*Lupinus* spp.) and the common ancestor of four other tribes, Phaseoleae (*Glycine*, *Psophocarpus*, and *Phaseolus* spp.), Trifolieae (*Medicago* spp.), Vicieae (*Pisum* and *Vicia* spp.), and Robinieae (*Sesbania* spp.).

Moreover, these data also suggest that the partial sequence of *nodC* can be used for identification of rhizobia at least to the species level. The classic molecular phylogenetic studies of

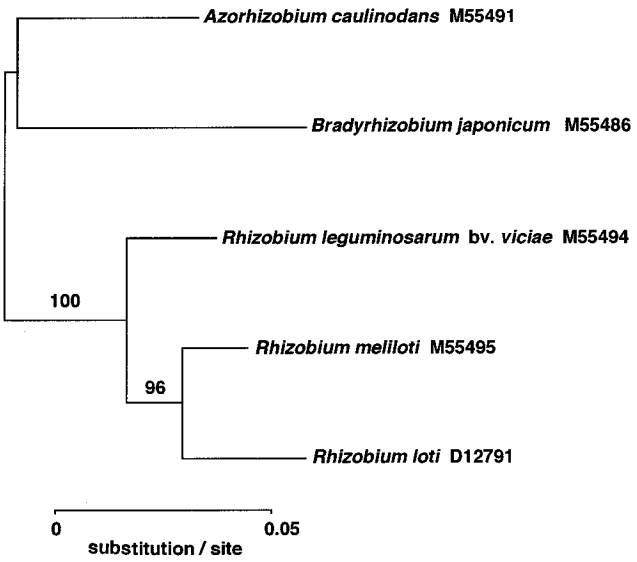



FIG. 4. Phylogenetic tree for 16S rRNA of five representative rhizobium strains. All five known 16S rRNA sequences were extracted from the GenBank, EMBL, and DDBJ DNA databases by using the ODEN system (8). The sequences were then aligned with each other by visual inspection to maximize the sequence similarity. In estimating evolutionary distances, we discarded regions common to the five sequences where deletions are observed even for one sequence. Omission of these regions left us 1,229 comparable nucleotide sites. To estimate evolutionary distances between nucleotide sequences, we used the method of Tajima and Nei (23). By using these estimates, molecular evolutionary trees were constructed from the resulting distance matrix by the NJ method (18) with the MEGA package (11). The database accession numbers are indicated after the bacterial names. The percentages of 10,000 bootstrap resamplings that support each topological element are indicated (4).

rhizobia have used 16S rRNA sequences (16, 24, 26). However, it is important to use other molecules at the same time, since conflicting molecular phylogenies may arise when lateral transfer among rhizobia is widespread. Further work on the molecular phylogenetics of rhizobia might require not only the identification of 16S rRNA molecules but also that of common nodulation genes such as *nodC*. In general, the degree of conservation is critical to resolving the phylogenetic problem at hand. For 16S rRNA sequences, it is reported that a partial sequence of 16S rRNA cannot be used for identification of rhizobia to the species level because of the high degree of conservation of the molecule (16). Moreover, the phylogeny of 16S rRNA molecules of rhizobia does not correspond with that of host plants. Therefore, the information on 16S rRNA alone might not be appropriate for identification and classification of rhizobia. The data presented here might form a framework that will aid in a reclassification of rhizobia. Furthermore, our data also suggest that phylogenetic analysis of common nodulation genes may be a significant way of assessing phylogenetic relationships in host plants.

Further phylogenetic analysis of more *nodC* sequences will be needed in order to elucidate the coevolution in more detail and is now in progress. Anyway, this molecular evolutionary study of the plasmid-legume symbiosis might provide insight into the evolution of symbiosis and might be important for increasing crop plant productivity by encouraging beneficial plant-microbe interactions.

**Nucleotide sequence accession numbers.** Sequences from the clones reported here have been submitted to the DDBJ nucleotide sequence database under accession numbers D28956 through 28965.

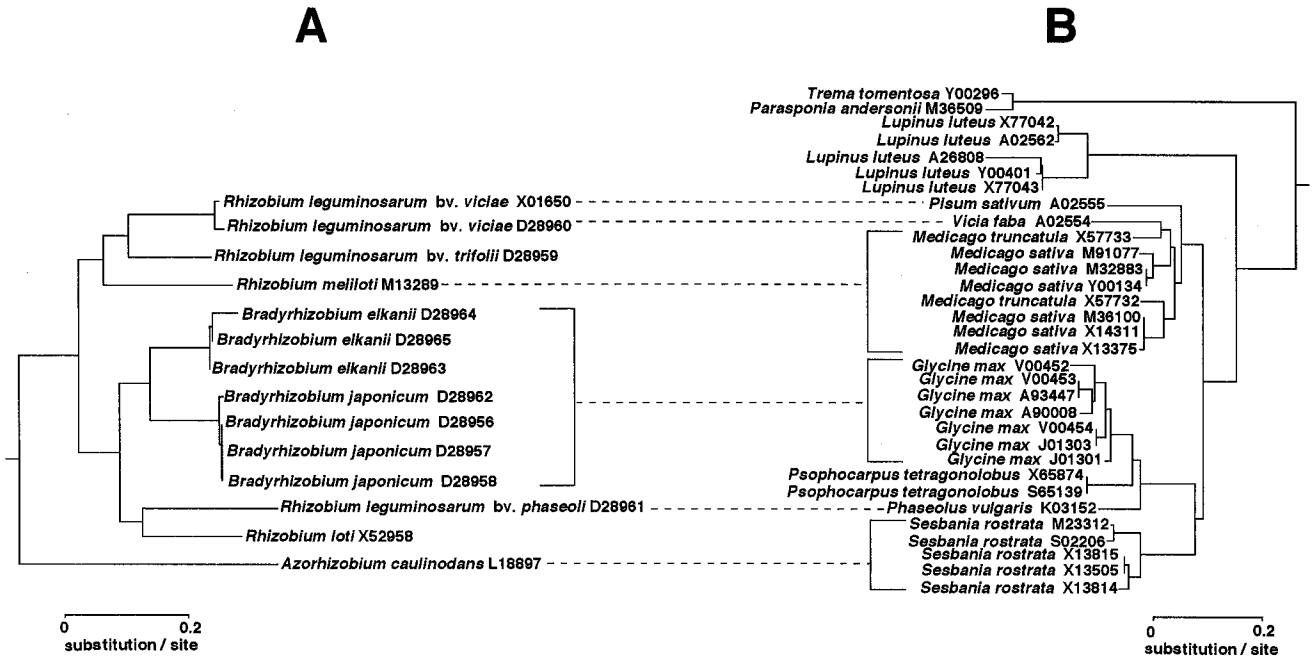



FIG. 5. Phylogenies of NodC protein sequences (A) and of leghemoglobin in host plants (B). The complete amino acid sequences of all 32 known leghemoglobins were extracted from the GenBank, EMBL, and DDBJ DNA databases and the PIR protein database by using the ODEN system (8). The complete leghemoglobin amino acid sequences were then aligned with each other by visual inspection to maximize the sequence similarity. In estimating evolutionary distances, we discarded the regions common to the 32 sequences where deletions are observed even for one sequence. Omission of these regions left us 134 comparable amino acid sites. To estimate evolutionary distances between amino acid sequences, we used Kimura's method (10). By using these estimates, molecular evolutionary trees were constructed from the resulting distance matrix by the NJ method (18) with the ODEN system (8). The database accession numbers are indicated after the bacterial and plant names.

We thank Y. Ina, Institute of Genetics, for the sequence analysis program and kind advice; T. Kajita and H. Ohashi, Tohoku University, and T. Mikami, Hokkaido University, for valuable advice about legume phylogeny; and P. V. Berkum for kindly providing bacterial strains from the USDA National *Rhizobium* Culture Collection.

## REFERENCES

1. **Bulawa, C. E., and W. Wasco.** 1991. Chitin and nodulation. Nature (London) **353:**710.
2. **Clarke, H. R. G., J. A. Leigh, and C. J. Douglas.** 1992. Molecular signals in the interactions between plants and microbes. Cell **71:**191–199.
3. **Doyle, J. J., M. Lavin, and A. Bruneau.** 1992. Contributions of molecular data to papilinoid legume systematics, p. 223–251. *In* D. E. Soltis, P. S. Soltis, and J. J. Doyle (ed.), Molecular systematics of plants. Chapman & Hall, London.
4. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
5. **Fisher, R. F., and S. R. Long.** 1992. *Rhizobium*-plant signal exchange. Nature (London) **357:**655–660.
6. **Futuyma, D. J., and M. Slatkin.** 1983. Introduction, p. 1–13. *In* D. J. Futuyma and M. Slatkin (ed.), Coevolution. Sinauer Associates Inc., Sunderland, Mass.
7. **Geremia, R. A., P. Mergaert, D. Geelen, M. V. Montagu, and M. Holsters.** 1994. The NodC protein of *Azorhizobium caulinodans* is an *N*-acetylglucosaminyltransferase. Proc. Natl. Acad. Sci. USA **91:**2669–2673.
8. **Ina, Y.** 1991. ODEN: molecular evolutionary analysis system for DNA and amino acid sequences, version 1.1.1. DNA Research Center, National Institute of Genetics, Mishima, Japan.
9. **John, M., J. Schmidt, U. Wieneke, H.-D. Krussmann, and J. Schell.** 1988. Transmembrane orientation and receptor-like structure of the Rhizobium meliloti common nodulation protein NodC. EMBO J. **7:**583–588.
10. **Kimura, M.** 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.
11. **Kumar, S., K. Tamura, and M. Nei.** 1993. MEGA: molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park.
12. **Moyle, W. R., R. K. Campbell, R. V. Myers, M. P. Bernard, Y. Han, and X. Wang.** 1994. Co-evolution of ligand-receptor pairs. Nature (London) **368:** 251–255.
13. **Nei, M.** 1975. Molecular population genetics and evolution. North-Holland, Amsterdam.
14. **Nei, M.** 1991. Relative efficiencies of different tree making methods for molecular data, p. 90–128. *In* M. M. Miyamoto and J. L. Cracraft (ed.), Recent advances in phylogenetic studies of DNA sequences. Oxford University Press, Oxford.
15. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
16. **Oyaizu, H., S. Matsumoto, K. Minamisawa, and T. Gamou.** 1993. Distribution of rhizobia in leguminous plants surveyed by phylogenetic identification. J. Gen. Appl. Microbiol. **39:**339–354.
17. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Ehrlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
18. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
19. **Schlaman, H. R. M., R. J. H. Okker, and B. J. J. Lugtenberg.** 1992. Regulation of nodulation gene expression by NodD in rhizobia. J. Bacteriol. **174:**5177–5182.
20. **Shadan, F. F., and L. P. Villarreal.** 1993. Co-evolution of persistently infecting small DNA viruses and their hosts linked to host-interactive regulatory domains. Proc. Natl. Acad. Sci. USA **90:**4117–4121.
21. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy. Freeman, San Francisco.
22. **Sprent, J. I.** 1994. Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? Plant Soil **161:**1–10.
23. **Tajima, F., and M. Nei.** 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1:**269–285.
24. **Yanagi, M., and K. Yamasato.** 1993. Phylogenetic analysis of the family Rhizobiaceae and related bacteria by sequencing of 16S rRNA gene using PCR and DNA sequence. FEMS Microbiol. Lett. **107:**115–120.
25. **Young, J. P. W.** 1993. Molecular phylogeny of rhizobia and their relatives, p. 587–592. *In* R. Palacios, J. Mora, and W. E. Newton (ed.), New horizons in nitrogen fixation. Kluwer Academic Publishers, London.
26. **Young, J. P. W., H. L. Downer, and B. D. Eardly.** 1991. Phylogeny of the phototrophic rhizobium strain BTAi1 by polymerase chain reaction-based sequencing of a 16S rRNA gene segment. J. Bacteriol. **173:**2271–2277.
27. **Young, J. P. W., and A. W. B. Johnston.** 1989. The evolution of specificity in the legume-rhizobium symbiosis. Trends Ecol. Evol. **4:**341–349.