Molecular Phylogeny of the Symbiotic Actinomycetes of the Genus *Frankia* Matches Host-Plant Infection Processes¹,²

Benoit Cournoyer, * Manolo Gouy, † and Philippe Normand*

*Laboratoire d'Ecologie Microbienne du Sol, and †Laboratoire de Biométrie, Génétique et Biologie des Populations, Université Claude-Bernard Lyon 1

Nucleotide sequences of approximately 213 bp of the *nifH*-D intergene and the beginning of *nifD* were determined for symbiotic *Frankia* isolates from the major host-infectivity groups. This region of the *nif* operon is variable enough to classify most infective *Frankia* strains at the species level. Phylogenetic inferences from these sequences are in agreement with the 16S rRNA-derived phylogeny of the genus and, thus, are in favor of an intrageneric evolution of *nif* genes by orthology. Phylogenetic lineages derived from combined *nifH*-D intergene and partial *nifD* and 16S rRNA sequences are supported for at least 93% of bootstrap replicates and are useful for investigating evolutionary relationships of the genus and symbiotic properties of this microorganism. The genus *Frankia* is divided into two major phylogenetic clusters that match with the separation of species according to the mechanism of infection of actinorhizal plants. One cluster groups species strictly adapted to the mechanism of root hair infection (RHI), and the other groups species adapted to the mechanism of direct intercellular penetration. In the RHI cluster, the species infective on *Casuarina* plants appears to have emerged from strains infective on *Alnus*. The concordance between the symbiotic properties and the molecular phylogeny of *Frankia* strains indicates a major role for the host plant in the evolution and speciation of the genus *Frankia*.

Introduction

The genus *Frankia* contains filamentous Gram-positive bacteria of the high-G+C% phylogenetic subdivision and is part of the Frankiaceae family (Woese 1987; Hahn et al. 1989). Species of *Frankia* have been described on the basis of the method of total DNA/DNA hybridization (Fernandez et al. 1989). Species defined by this method, which is considered by Wayne et al. (1987) to be the most objective means for delineating bacterial species, are termed “genomic species.” The interactions between actinomycetes of the genus *Frankia* and about 25 different genera of woody dicots result in the development of nitrogen-fixing root nodules and the transfer of fixed nitrogen from the microsymbiont to the plant.

Host-specific responses divide *Frankia* isolates into three major host-infectivity groups (Huang et al. 1985; Baker 1987): (A) strains infective on *Alnus* species and

1. Key words: *nifD*, nitrogenase, molecular phylogeny, symbiosis, actinomycete, *Frankia* species.

2. The publication of this paper was delayed by six months through no fault of the authors. *Molecular Biology and Evolution* accepts full responsibility for the delay. We apologize to the authors and to our readers for that unreasonable delay.

Address for correspondence and reprints: Benoit Cournoyer, University of the West of England, Faculty of Applied Sciences, Biological Sciences Department, Coldharbour Lane, Bristol BS16 1QY, United Kingdom.
Myrica species, (B) strains infective on Casuarina species and Myrica species, and (C) strains strictly infective on Elaeagnus species or infective on Elaeagnus species and Myrica species or Alnus species. Frankia isolates are also divided into two major groups according to the adaptation of strains to the host plant–specific responses to invasion. One group is adapted for a root hair infection (RHI) process and contains strains representing host-infectivity groups A, B, and, in some cases, C. For RHI, after inoculation with Frankia, plant root hairs become deformed and are invaded by the actinomycete. Frankia reaches intracellularly the cortical cells of the nodule primordium, and a nodule lobe develops by the enlargement of these infected cortical cells (Callaham et al. 1979). The other group of strains is adapted for a direct intercellular penetration (IP) process and is represented by the host-infectivity group C. For IP, Frankia colonizes the root surface and penetrates the root through the middle lamellae between epidermal cells (Miller and Baker 1985). Root hairs are not involved in this process. Frankia gets to the nodule primordium by intercellular colonization of the root cortex. Some Frankia strains are capable of employing both RHI and IP colonization mechanisms. The host plant determines which one is used (Miller and Baker 1986). Estimation of evolutionary distances between the various groups described above might give insights on the evolution of the genus and the symbiotic properties of the strains.

The phylogenetic relationships established recently for isolates of the genus Frankia were based on partial 16S rRNA gene sequences (Nazaret et al. 1991). The quantity of informative sites, the level of significance of the lineages, and whether these would be representative of the whole genome led us to reexamine this phylogeny. In the present paper, we report on DNA sequences of the nifH-D intergene and the beginning of nifD from genomic species of the major host-infectivity groups, for the phylogenetic analysis of the genus Frankia. The nifD gene encodes an apoprotein of the dinitrogenase complex involved in the reduction of atmospheric nitrogen into ammonia and is essential for the nitrogen-fixing activity of Frankia strains in symbiosis. By comparing the recently reported nifD gene from F. alni strain Ar13 (Normand et al. 1992) with nifD sequences from other nitrogen fixers, we have been able to define well-conserved oligonucleotide sequences for polymerase chain reaction (PCR) amplification and sequencing of these regions.

Material and Methods
Frankia Strains and DNA Extraction

Frankia strains from the three major plant-infectivity groups were used in the present study (table 1). Strains of the Alnus-infectivity group were grown in Frankia-Tween (FTW) medium (Simonet et al. 1985). Strains of the Elaeagnus-infectivity group were grown in FTW medium without Tween 80. The strain of the Casuarina-infectivity group was grown in buffer of an ammonium chloride sodium propionate (Murry et al. 1984). Total genomic DNA of each strain was extracted by following the procedure described by Simonet et al. (1984).

PCR Amplification and Sequencing

PCR was used to amplify segments of double-stranded (ds) DNA delineated by oligonucleotides FGPH750 to FGPD224' (fig. 1). These segments run from nucleotide (nt) 750 at the 3' end of the nifH coding region to nt 224 of the nifD cistron. Both of these primers are complementary to highly conserved regions within nifH and nifD coding regions and have been successfully used for amplification by PCR. Mul-
multiple alignment of \textit{nifD} sequences showed FGPD224' to be a well-conserved oligonucleotide sequence among nitrogen-fixing bacteria (fig. 1) that corresponds to the amino acid motif RGCAYAG. The oligonucleotide sequence FGPH750 was previously used for \textit{Frankia} genus-specific characterization by PCR (Simonet et al. 1991). The oligonucleotide sequence FGPD161' is well conserved among \textit{nifD} genes, but gaps were introduced to match this sequence with the homologous sequence of \textit{Azotobacter vinelandii} (fig. 1).

PCR (50 μl) was carried out according to the procedure described by Simonet et al. (1991). Thirty-five PCR cycles were carried out on an automated thermocycler (Trio-Thermoblock; Biometra) with the following parameters: 94°C for 1 min, 57°C for 1 min, and 72°C for 1 min. Detection of PCR amplifications and controls was done according to the method of Simonet et al. (1991).

A method for the direct sequencing of amplified dsDNA was used (Winship 1989). The fragments were sequenced in both directions by using primers FGPH750, FGPD224', and FGPD161'.

**DNA Sequence Analysis**

**Evolutionary Distances and Phylogenetic Trees**

DNA sequences were aligned using the multiple-alignment clustal algorithm (Higgins and Sharp 1988) with some manual refinements in the noncoding regions. Sites involving gaps were excluded from all analyses. Evolutionary distances and their standard deviations (SD; the square root of the variance) were computed using Jukes and Cantor's equations (Nei 1987, pp. 65–66). Phylogenetic trees were inferred using
<table>
<thead>
<tr>
<th>species</th>
<th>FGPH750</th>
<th>FGPD161</th>
<th>FGPD224</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Frankia alni</em></td>
<td>GAAGACGATCCCTACCCCGA</td>
<td>GTCAAGTCCAACATCAAATGTC</td>
<td>CTACGCCGGTTCCAAGGGCG</td>
</tr>
<tr>
<td><em>Anabaena sp.</em></td>
<td>GCTCACCATTCCTACACCAA</td>
<td>GTTAAGTCTACATCAAATATC</td>
<td>TATGCGAGGTCTAAGGGTG</td>
</tr>
<tr>
<td><em>Azospirillum brasilense</em></td>
<td>GGGCACCACATCCCGACCCCGA</td>
<td>GTCAAGTCCGAACATCAAATGTC</td>
<td>CTACGCCGGTTCCAAGGGCG</td>
</tr>
<tr>
<td><em>Azotobacter vinlandii</em></td>
<td>GCTGGTCATCCCGAACCCGA</td>
<td>ATCCAGTCCAACATCAAATC</td>
<td>CTACGCCGGTTCCAAGGGCG</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em></td>
<td>GAAAAGTGGGCACGCGGCGCT</td>
<td>ATATTCCTTAACCGCAAATGTC</td>
<td>CTACGCCGGTTCCAAGGGCG</td>
</tr>
<tr>
<td><em>Rhizobium japonicum</em></td>
<td>GGGCATCATTCCGACCGCGGA</td>
<td>GTGAAATGTTCCATCAAAATC</td>
<td>CTATGCAGGGTGCAAGGGCG</td>
</tr>
<tr>
<td><em>Thiobacillus ferrooxidans</em></td>
<td>CGCGCTATCCCGACCGCGGA</td>
<td>GTCAAGTCCCAACATCAAATC</td>
<td>CTACGCCGGTTCCAAGGGCG</td>
</tr>
</tbody>
</table>

**Fig. 1.**—Sequences of the PCR amplification and/or sequencing primers aligned with homologous regions of *nifH* and *nifD* genes from other bacteria. The sequence of *Frankia alni* was used to define primers. For PCR amplification and sequencing, the antisense sequences of FGPD161 and FGPD224 were used and named FGPD161' and FGPD224', respectively. Primer FGPH750 was defined by Simonet et al. (1991). Sequence sources are *Frankia*, Normand and Bousquet (1989) and Normand et al. (1992); *Anabaena*, Mevarech et al. (1980) and Lammers and Haselkorn (1983); *Azospirillum*, GenBank accession number M64344; *Azotobacter*, Brigle et al. (1985); Klebsiella, Arnold et al. (1988); *Rhizobium*, Kaluza et al. (1983) and Kaluza and Hennecke (1984); and *Thiobacillus*, Pretorius et al. (1987) and Rawlings (1988).

Relative-Rate Tests

The relative-rate test allows one to compare the rate of evolution of two lineages, 1 and 2, since divergence from their last common ancestor by employing a homologous reference sequence 3 that is evolutionarily external to sequences 1 and 2. If \( d_{ij} \) denotes the evolutionary distance between sequences \( i \) and \( j \), the difference in evolution between lineages 1 and 2 can be measured by \( d_{13} - d_{23} \). The relative-rate test computes the variance of this quantity. This test was used according to the method of Wu and Li (1985) and was simplified for use of Jukes and Cantor’s one-parameter model of base substitution. If it is assumed that \( d_{13} - d_{23} \) has a Gaussian distribution, lineages 1 and 2 are taken to have evolved with the same rate since divergence from their last common ancestor, when the absolute value of \( d_{13} - d_{23} \) is less than twice its SD.

Dating of an Internal Node in a Phylogenetic Tree

Let \( I \) be the internal node to be dated, and let the tree tips descending from node \( I \) be 1, 2, \ldots, \( m \). The average distance from node to tips \( d_{r,[1 \ldots m]} \) can be computed from the tree by adding the lengths of branches connecting node \( I \) to the tips and averaging these lengths over the \( m \) tips. If all \( m \) lineages evolved with the same rate of nucleotide substitution, \( r \), the date of the internal node \( I \) is

\[
T(I) = \frac{d_{r,[1 \ldots m]}}{r}.
\]

The variance of this date estimation is given by

\[
\text{Var}[T(I)] = \frac{1}{r^2} \text{Var}(d_{r,[1 \ldots m]})
\]
in the absence of known estimation of the accuracy of parameter \( r \). The variance of the node-to-tips evolutionary distance \( d_{r,11} \ldots m \) can be estimated as follows: With \( p \), frequency of nucleotide differences between two sequences, and \( L \), sequence length, the evolutionary distance \( d \) between sequences and its variance \( \text{Var}(d) \) are given by Jukes and Cantor's equations (Nei 1987, pp. 65–66)

\[
d = -\frac{3}{4} \log \left( 1 - \frac{4}{3} p \right) ;
\]

and

\[
\text{Var}(d) = \frac{p(1-p)}{\left(1 - \frac{4}{3} p \right)^2 L}.
\]

Inverting equation (1) gives

\[
p = \frac{3}{4} \left[1 - \exp \left(-\frac{4}{3} d\right) \right].
\]

Thus \( \text{Var}(d) \) and \( d \) are functionally related as shown by putting equation (3) into equation (2). Wu and Li (1985) have applied this relationship between \( d \) and \( \text{Var}(d) \) to estimate the variance of the distance between a tree tip and an internal node. This relationship is used here to compute \( \text{Var}(d_{r,11} \ldots m) \) from \( d_{r,11} \ldots m \).

Results
PCR Amplification and DNA Sequences of \( nifHD \) Segments

All strains listed in table 1, except Ar13 whose complete \( nifD \) sequence was reported elsewhere (Normand et al. 1992), were studied. PCR amplification of total genomic DNA from these \( Frankia \) strains by using primers FGPH750 and FGPD224' produced a single fragment [from 375 bp in length, for strains HRN 18a, EaII-41 (also called "E41"), and M2, to 400 bp, for ARgP5 \( ^{AG} \)]. DNA sequences obtained from these amplified dsDNA segments (fig. 2) were read in both directions, beginning at the stop codon of \( nifH \) (position 961, according to Normand and Bousquet 1989) and ending at position 142 of \( F. \) \( alni \) strain Ar13 \( nifD \) numbering (Normand et al. 1992). Length of these sequences varied according to nt indels in the intergenic spacer between \( nifH \) and \( nifD \). The \( nifH-D \) intergene varies in length from the observed 43 nt of HRN18a (genomic species 7) to the 68 nt of strain ARgP5 \( ^{AG} \) (genomic species 3). Most genomic species have their own specific \( nifH-D \) intergene sequence. However, genomic species 4 (HRX40la) and 5 (Eal-12) are undifferentiable. The closely related strains ACN1 \( ^{AG} \), Ar13, and ACN14a are differentiated into two groups according to the base (cytosine or thymine) at position 47 of the \( nifH-D \) intergene. The \( nifH-D \) intergene and \( nifD \) sequence from strain Ar13 (Normand et al. 1992) are identical to the sequence from ACN14a. Overall, among all those sequences, 45 variable alignable sites were recorded. The putative tetranucleotide ribosome binding site for translation of \( nifD \) is shown in figure 2. Of the 10 \( Frankia \) strains whose \( nifH-D \) intergene and beginning of \( nifD \) sequences are reported here, all but EaII-41 had a partial 16S rRNA gene sequence reported elsewhere (Nazaret et al. 1991).
Fig. 2.—Alignment of the nucleotide sequences of the nif H-D intergene and beginning of nif D from 11 Frankia strains belonging to the three major host-infectivity groups. Sequences begin at the nif H stop codon. Dashes denote gaps introduced by the multiple alignment. The sequence of strain ArI-13 is from Normand et al. (1992). Nucleotides from position 4–38 were excluded from phylogenetic analysis because alignment of this region is uncertain. For strain Eal-12, the putative tetranucleotide ribosome binding site for nif D translation is underlined, and the start codon is indicated by three asterisks.
Phylogeny of Frankia

Phylogenetic relationships between Frankia strains were established by the NJ algorithm using pairwise evolutionary distances and by the MP method. The significance of internal branches of NJ trees was assessed by applying the NJ algorithm to 1,000 bootstrap replicates. Both tree-building methods were used with the nifH-D intergene and partial nifD sequences (data not shown), the 16S rRNA fragments (data not shown), and the combined nif and 16S rRNA sequence data (fig. 3). Phylogenetic relationships between strains common to the three sequence sets are identical in the three NJ analyses. The bootstrap analysis of the phylogenetic tree for the nifH-D intergene and partial nifD sequences shows that these sequences alone resolve only a small fraction of the tree (e.g., the HRN18a,EUN1f,EaII-41 cluster) in a statistically significant way. The MP analysis of these sequences identifies six equally parsimonious trees requiring 72 substitutions. One of these trees has the topology of the NJ tree. The clusters of the NJ phylogenetic tree supported by 87% or more of bootstrap replicates are shared by all six most parsimonious trees. Thus MP and NJ analyses give completely congruent results for the nifH-D data. Rooting this NJ phylogenetic tree with the closest known outgroup sequence, that of Anabaena (Lammers and Haselkorn 1983), was attempted but shows a high level of uncertainty, because the outgroup is too evolutionarily distant.

![Phylogenetic tree](image)

**Fig. 3.**—Phylogenetic tree of symbiotic Frankia, using the combined data set, nif, and 16S rRNA, but rooted and dated according to 16S rRNA alone. Phylogenetic relatedness among these genomic species of the genus are based on the NJ method using nucleic acid pairwise distances. Horizontal distances are proportional to evolutionary divergences expressed in substitutions per 100 sites; vertical separations are for clarity only. Bootstrap $P$ values of each cluster are indicated in ovals. nifH-D intergene and beginning of nifD sequences are from fig. 2, and partial 16S rRNA sequences are from Nazaret et al. (1991) and Cournoyer and Normand (1992). Four hundred forty-five homologous sites were used. The tree is rooted by reference to 16S rRNA phylogenetic analysis alone, which contains a known outgroup, Actinooplanes brasiliensis (Cournoyer 1993, p. 61). The root-containing branch was divided into two parts, of length 1.6 and 0.3, proportionate to the size of corresponding branches in the 16S rRNA analysis. The Frankia strain EaII-41 was sequenced only at its nif locus and was thus positioned in this phylogenetic tree by maintaining the same proportion in branch length as that observed in the nif-alone tree (indicated by a dotted line). The given estimated date of the radiation of symbiotic Frankia is based on analysis of 16S rRNA data only.
Nevertheless, the 16S rRNA tree can be rooted using the homologous sequence of *Actinoplanes brasilienensis* (Cournoyer 1993, p. 61), a closely related actinomycete (Woese 1987). MP analysis of the partial 16S rRNA sequence data identifies 25 equally parsimonious trees requiring 49 substitutions. One of them has the topology of figure 3. The two monophyletic clusters of the 16S rRNA tree supported by more than 95% of bootstrap replicates, i.e., the four *Elaeagnus*-infective strains and the ARgP5 AG–AVN17o cluster, are also present in all 25 most parsimonious trees (clusters Eal-12–HRX401a and Arl3-ACN1 AG–ACN14a are trivial because they contain identical sequences). Thus NJ and MP analyses of the 16S rRNA data are congruent in indicating that *Elaeagnus*-infective strains form a monophyletic cluster.

Because the phylogenetic trees for the *nif* H-D intergene and partial *nif* D and for the partial 16S rRNA sequences do not differ in topology, one may combine the two molecules to obtain better resolution of the evolutionary divergence of *Frankia* strains. The NJ tree of combined *nif* H-D and 16S rRNA sequence data is supported by more than 93% of bootstrap replicates in most of its branchings (fig. 3). MP analysis of these sequences identifies three equally parsimonious trees requiring 92 substitutions (445 sites analyzed, 59 varied, 41 cladistically informative) that differ only in the relationships among strains Arl3, ACN14a, and ACN1 AG and that are identical elsewhere to the NJ tree. Thus NJ and MP analyses applied to the combined data set yield the same evolutionary relationships for *Frankia* strains: strains from the *Elaeagnus*-infectivity group form a monophyletic cluster; strains of *F. alni* (ACN1 AG and Arl3) and strain ACN14a are more closely related to the *Casuarina*-infective strain (M2) than to other genomic species of the *Alnus*-infectivity group.

Evolutionary distances between *Frankia* strains for the combined *nif* H-D–16S rRNA data are all lower than 0.09 substitutions/site (table 2). The contribution of correction for multiple hits in pairwise distances never exceeds 0.005. NJ analysis of the *nif* H-D–16S rRNA data without a distance-correction factor gives results identical to those found by using a correction factor.

### Table 2

<table>
<thead>
<tr>
<th>Eal-12</th>
<th>HRN18a</th>
<th>FIUN1f</th>
<th>ARgP5 AG</th>
<th>AVN17o</th>
<th>Arl3</th>
<th>ACN1 AG</th>
<th>M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eal-12</td>
<td>5.1 ± 1.1</td>
<td>2.5 ± 0.8</td>
<td>8.6 ± 1.5</td>
<td>8.3 ± 1.4</td>
<td>3.7 ± 0.9</td>
<td>1.0 ± 1.3</td>
<td>1.0 ± 1.3</td>
</tr>
<tr>
<td>HRN18a</td>
<td>6.8 ± 1.3</td>
<td>7.8 ± 1.4</td>
<td>8.6 ± 1.5</td>
<td>8.3 ± 1.4</td>
<td>3.7 ± 0.9</td>
<td>1.0 ± 1.3</td>
<td>1.0 ± 1.3</td>
</tr>
<tr>
<td>FIUN1f</td>
<td>7.8 ± 1.4</td>
<td>8.6 ± 1.4</td>
<td>8.6 ± 1.4</td>
<td>6.1 ± 1.2</td>
<td>6.6 ± 1.3</td>
<td>6.6 ± 1.3</td>
<td>0.3 ± 0.2</td>
</tr>
<tr>
<td>ARgP5 AG</td>
<td>2.8 ± 1.2</td>
<td>8.5 ± 1.4</td>
<td>7.6 ± 1.4</td>
<td>6.1 ± 1.2</td>
<td>6.6 ± 1.3</td>
<td>6.6 ± 1.3</td>
<td>0.3 ± 0.2</td>
</tr>
<tr>
<td>AVN17o</td>
<td>6.1 ± 1.2</td>
<td>8.3 ± 1.4</td>
<td>7.8 ± 1.4</td>
<td>6.1 ± 1.2</td>
<td>6.8 ± 1.3</td>
<td>6.8 ± 1.3</td>
<td>0.2 ± 0.2</td>
</tr>
<tr>
<td>Arl3</td>
<td>7.8 ± 1.4</td>
<td>8.3 ± 1.4</td>
<td>7.8 ± 1.4</td>
<td>7.3 ± 1.3</td>
<td>7.6 ± 1.4</td>
<td>4.9 ± 1.1</td>
<td>5.1 ± 1.1</td>
</tr>
</tbody>
</table>

**Note:**—Evolutionary distances (below the diagonal) and relative-rate tests between strains on the same side of the root of fig. 3 (above the diagonal) were computed with Jukes and Cantor's equations and are given in substitutions per 100 sites. Strains HRX401a and ACN14a are not shown because their sequences are identical to those of strains Eal-12 and Arl3, respectively.
common ancestor. This was done with the relative-rate test (Wu and Li 1985) on the combined \textit{nif} H-D–16S rRNA data of figure 3. As a matter of principle, one should always use the closest outgroup to the pair of taxa being tested to conduct a relative-rate test. Thus the relative-rate test was applied here in two phases. First, we compared the rate of evolution of any given pair of strains located on the same side of the root of figure 3 by using all most closely related strains as the outgroup. Second, we tested for equal average rates of evolution on the two sides of the root. Thus, when strains ARgP5\textsuperscript{AG} and AVN17o were considered, their average distances to strains ArI3, ACN1\textsuperscript{AG}, and M2 were used to conduct the test; similarly, for pair Eal-12–HRN18a, strains ARgP5\textsuperscript{AG}, AVN17o, ArI3, ACN1\textsuperscript{AG}, and M2 were simultaneously used as a composite outgroup. Results of these pairwise relative-rate tests are shown in table 2. In all cases, \(d_1 - d_2\) is less than twice its SD, which shows that there is no evidence for unequal rates of evolution within each side of the root. The average distances from the root to the tree tips are 4.1 ± 1.0 and 3.8 ± 0.9 substitutions/100 sites for the HRN18a- and M2-containing sides, respectively (for SD computation, see Material and Methods). Thus, \textit{Frankia} strains on both sides also appear to have evolved at similar rates.

\textbf{Comparison of the Rate of Evolution of the Studied 16S rRNA Gene Fragment with That of the Complete Gene}

Complete 16S rRNA sequences from eight high-G+C, Gram-positive bacteria (\textit{Actinomycetes pyogenes}, GenBank accession no. M29552; \textit{Arthrobacter globiformis}, M23411; \textit{Mycobacterium bovis}, M20940; \textit{Streptomyces coelicolor}, Y00411; \textit{Frankia} sp., M55343; \textit{Propionibacterium freudenreichii}, X53217; \textit{Saccharomonospora viridis}, X54286; and \textit{Terrabacter tumescens}, X53215) were aligned. Evolutionary distances between sequence pairs corrected by the two-parameter method (Nei 1987, p. 67) were computed from complete sequences and from the fragment homologous to that sequenced in \textit{Frankia} strains. Fragment-based distances were plotted as a function of complete sequence-based distances for the 28 sequence pairs. The center of the plot (data not shown) lay at coordinates (0.119, 0.100), showing that the ratio of the evolutionary rate of the fragment to that of the complete gene is 0.84. Thus, this fragment evolves at a rate of approximately 0.008 substitutions/site/50 Myr, instead of the reported 1%/50 Myr for the complete molecule (Ochman and Wilson 1987). The average distance between the root of the \textit{Frankia} radiation and the tree tips in the 16S rRNA NJ trcc is 0.021 ± 0.0089 substitutions/site (for computation of this distance and its SD, see Material and Methods). Thus, the age of this radiation can be dated to 50(0.021/0.008) ± 50(0.0089/0.008) = 131 ± 56 Myr.

\textbf{Discussion}

The nitrogen-fixing actinomycetes of the genus \textit{Frankia} have been known for several years to be in symbiosis with pioneer woody dicots. However, isolates that can neither infect plants nor fix nitrogen but that are closely related to the genus \textit{Frankia} according to fatty acid profiles have recently been described by Mirza et al. (1991). This suggests the presence of a hidden diversity within the genus and questions the importance of the symbiosis to the evolution and speciation of \textit{Frankia}. In the present paper, we determine phylogenetic relationships among isolates from the major host-infectivity groups to investigate the relative importance of the host plant to the evolution of the genus.

The phylogeny of the genus \textit{Frankia} is assessed from the \textit{nif} H-D intergene and
the beginning of \textit{nif} D and from partial sequences from the 16S rRNA gene. There is a specific \textit{nif} H-D intergene and beginning of \textit{nif} D sequence for each genomic species of the \textit{Alnus}- and \textit{Casuarina}-infectivity groups. The genomic species of strain ACN14a has not been determined, but the sequence data presented here strongly suggest that it belongs to species 1, \textit{F. alni}. In contrast to the partial 16S rRNA genes (Nazaret et al. 1991), the \textit{nif} H-D intergene and beginning of \textit{nif} D sequences are more heterogeneous among genomic species from the \textit{Elaeagnus}-infectivity group than those from the \textit{Alnus}-infectivity group. Therefore, genomic species of the \textit{Elaeagnus}-infectivity group are more easily differentiated by using the \textit{nif} H-D intergene and beginning of \textit{nif} D than by using the partial 16S rRNA genes. Genomic species 4 (EaII-12) and 5 (HRX40Ia) remain undifferentiable from each other with both sequence sets.

Phylogenetic relationships between \textit{Frankia} strains inferred from \textit{nif}ID, 16S rRNA, and the combined sequence data (fig. 3) are identical. Thus, these \textit{nif} genes have evolved in the genus \textit{Frankia} in parallel with other genes and without lateral transfer events. In \textit{Rhizobium}, another nitrogen-fixing symbiotic bacterium, \textit{nif} genes are located on plasmids, and there is evidence for lateral transfer of these plasmids between closely related species (Broughton et al. 1987; Eardly et al. 1992). In \textit{Frankia}, \textit{nif} genes have been located on a large plasmid of the strain ARgP5\text{AG} (Simonet et al. 1986), but it remains unknown whether \textit{nif} genes of other strains are located on the bacterial chromosome or on large plasmids. All three trees reveal a major separation between the \textit{Elaeagnus}-infectivity group and the \textit{Alnus}- and \textit{Casuarina}-infectivity groups. With the partial 16S rRNA gene sequences, the location of the tree root can be sought. In light of the limited length of the sequenced fragment, the root location cannot be unambiguously determined but can be excluded from within the group of \textit{Elaeagnus}-infective strains.

Phylogenetic relationships within the genus \textit{Frankia} are best assessed with the combined data set (fig. 3). Most, if not all, branchings of this tree are statistically significant, which indicates that the radiation of the genus \textit{Frankia} has been reliably analyzed. Moreover, the same phylogenetic relationships are found when data are analyzed with Kimura’s two-parameter distance correction. Because interstrain distances are moderate (<9%), the effect of distance correction is very low (<0.5%). \textit{Elaeagnus}-infective strains form a monophyletic cluster. This cluster is further divided between genomic species 6 (EUN1f) and 7 (HRN18a) on one side and species 4 (EaII-12) and 5 (HRX40Ia) on the other. Strains of the \textit{Alnus}-infectivity group are divided into two parts. Genomic species 2 (AVN17o) and 3 (ARgP5\text{AG}) are separated from \textit{F. alni} (Ar13, ACN1g) and strain ACN14a. The \textit{Casuarina}-infective strain (M2) appears specifically related to \textit{F. alni} strains (supported by 93% of bootstrap replicates).

The phylogenetic relationships match the mode of host-plant infection: strains adapted for colonization by IP (i.e., those infective on \textit{Elaeagnus}) form a monophyletic cluster. Among them, some are adapted for colonization by both IP and RHI (EaII-41 and EUN1f) and are termed “flexible strains.” These two flexible strains, which infect both \textit{Alnus} and \textit{Elaeagnus}, are probably phylogenetically related, since they are clustered at a significant level on figure 3 (EaII-41 was sequenced only at its \textit{nif} locus). Strains used to evaluate phylogenetic relationships among genomic species strictly adapted for RHI are specialized in the infection of either \textit{Alnus} or \textit{Casuarina}. Genomic species of the \textit{Alnus}-infectivity group appear to have diverged prior to the emergence of the genomic species of the \textit{Casuarina}-infectivity group. The \textit{Casuarina}-infectivity group seems to have emerged from the species \textit{F. alni}. This is not surprising, because the host-plant genera \textit{Alnus} and \textit{Casuarina} belong to the same subclass, the Hama-
meliada (Cronquist 1988, pp. 292–298). Moreover, the first pollen occurrence of the Casuarinaceae family has been dated at ~60 Mya and is more recent than the appearance of the Betulaceae, to which Alnus species belong. This suggests a first adaptation of Frankia to the colonization of Alnus and a further specialization with the appearance of Casuarina plants.

Steady rates of base substitution for periods as long as 50–100 Myr have been reported for eubacterial 16S rRNA (Wilson et al. 1987). The partial gene sequences analyzed here are shown by the relative-rate test to have evolved with equal rate in all strains. The evolutionary radiation of the genus Frankia was therefore dated after calibration of the molecular clock reported for the complete 16S rRNA to the fragment used. The obtained age of 131 ± 56 Myr corresponds to the approximate date of appearance and diversification (~100 Mya) of the Betulaceae (Thomas and Spicer 1987, p. 309) and provides support for the major role of actinorhizal plants on the evolution of the genus. The SD of this estimate, which is quite large (56 Myr), takes into account the sampling error involved in estimating pairwise evolutionary distances but ignores any error associated with the time calibration of the tree, because such data are not available. The accuracy of the estimation could only be improved by increasing the length of the 16S rRNA sequences. The rough agreement between age of the Frankia radiation and age of diversification of the Betulaceae provides another key event, in the evolution of microbes, in support of the calibration of the rRNA molecular clock for prokaryotes.

The concordances established between the symbiotic properties and the evolution of the microorganism indicate an impact of the host plant on the evolution of Frankia, in at least two instances: (1) at the time of divergence between strains adapted for RHI and IP and (2) with the specialization of one or more strains of the Alnus-infectivity group into a Casuarina-infectivity group. Other speciations might not be related to the evolution of the symbioses. Frankia is not a strictly symbiotic microorganism, and adaptations under free-living conditions could have been determinantal for its evolution. However, its free-living status has received little attention. By using PCR, it is now possible to follow specific microorganisms in soils and sediments (Tsai and Olson 1992), and an evaluation of the distribution of the various genomic species of Frankia in soils may allow one to explain other aspects of the evolution of the genus.

Acknowledgments

B.C. thanks the government of Québec (Canada) for receipt of an FCAR award. We would like to thank M.-C. Grosjean and S. Nazaret for useful advice throughout this work. This work was partly supported by grant 603-1 from the INDO-French Centre for the Promotion of Advanced Research.

LITERATURE CITED


WHEELER, and D. A. PEKKY, eds. Symbiotic nitrogen fixation in the management of temperate forests. Forest Research Laboratory, Oregon State University, Corvallis.


Molecular Phylogeny of *Frankia*


ELIZABETH ZIMMER, reviewing editor

Received July 1, 1992; revision received July 21, 1993

Accepted July 21, 1993