Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences

Michael L. Clawson, Aaron Bourret, and David R. Benson

Abstract

Actinomycetes from the genus *Frankia* induce nitrogen-fixing root nodules on actinorhizal plants in the “core rosids” clade of eudicots. Reported here are nine partial *Frankia* 16S rRNA gene sequences including the first from host plants of the rosaceous genera *Cercocarpus* and *Chamaebatia*, 24 partial glutamine synthetase (GSI; *glnA*) sequences from *Frankia* in nodules of 17 of the 23 actinorhizal genera, and the partial *glnA* sequence of *Acidothermus cellulolyticus*. Phylogenetic analyses of combined *Frankia* 16S rDNA and *glnA* sequences indicate that infective strains belong to three major clades (I–III) and that Clade I strains consisting of unisolated symbionts from the Coriariaceae, Datiscaceae, Rosaceae, and *Ceanothus* of the Rhamnaceae are basal to the other clades. Clock-like mutation rates in *glnA* sequence alignments indicate that all three major *Frankia* clades diverged early during the emergence of eudicots in the Cretaceous period, and suggest that present-day symbioses are the result of an ancestral symbiosis that emerged before the divergence of extant actinorhizal plants.

Keywords: Actinorhizal; Root nodules; *Frankia*; Symbiosis; Evolution; Nitrogen-fixing

1. Introduction

Angiosperms that bear N₂-fixing root nodules occur in 10 families distributed among five of eight main lineages of the “core rosids” (Magallon et al., 1999). Two groups of bacteria participate in the symbioses. One group consists of diverse proteobacteria possessing the rhizobial nodulation genes that confer the ability to nodulate legumes of the Fabaceae and the single genus *Cercocarpus* from the Fabaceae and the single genus *Parasponia* from the Ulmaceae. The other group consists of high mol% G + C actinomycetes from the genus *Frankia* that nodulate plants from eight additional families shown in Table 1 (Benson and Silvester, 1993).

The restriction of nodulating ability to the core rosids has led to the hypothesis that the predisposition to form nitrogen-fixing root nodule symbioses emerged once during angiosperm evolution (Soltis et al., 1995). The diversity of plants infected by *Frankia* has led to the suggestion that actinorhizal plants expressed this predisposition for symbiosis on multiple, perhaps three to six, occasions during angiosperm evolution (Jeong et al., 1999; Roy and Bousquet, 1996; Swensen, 1996).

From the bacterial perspective, phylogenetic studies of effective (capable of nodulating plant roots and fixing nitrogen) *Frankia* strains based on 16S rDNA or *nifH* sequences have generally revealed three major clades (Benson and Clawson, 2000; Jeong et al., 1999; Nick et al., 1992; Normand et al., 1996; Wolters et al., 1997). *Frankia* strains in these clades can be defined by behavior in culture, by morphology, and by mode of infection (Benson and Silvester, 1993). Clades I have not been isolated in culture, infect hosts by intercellular penetration through the root surface (Berry and Sunell, 1990; Kohls et al., 1994; Liu and Berry, 1991; Miller and Baker, 1985, 1986; Valverde and Wall, 1999) and form primitive vesicles in symbiosis (Newcomb et al., 1987).
in culture and form multicellular lipid-enveloped structures termed vesicles in both culture and symbiosis. Clade II strains prefer to grow on organic acids, whereas Clade III strains grow on both organic acids and simple sugars (Benson and Silvester, 1993). Although relatively few plants have been studied, Clade II strains appear to infect their hosts by root hair infection and Clade III strains infect by either intercellular penetration or root hair infection depending on the plant being infected (Berry et al., 1986; Bosco et al., 1992; Callaham et al., 1996). Portions of the 16S rDNA, glnA genes were amplified from the same DNA preparations. Polymerase chain reaction (PCR) primers used for the 16S rRNA gene were fD1 (Weisburg et al., 1991) and rDB1 (Benson and Silvester, 1993). The glnA gene fragments were designated DB43 (5'-TTCTTTCATCCACGGC-3') and DB44 (5'-GGACCGGAAG-TAGCC-3') and yielded a 477 bp fragment encompassing roughly one-third of the gene.

2. Materials and methods

2.1. Bacterial strains

Table 2 lists the sources of sequences used in this study. Root nodules were collected from 17 of the 23 known actinorhizal plant genera and stored in 95% ethanol at −20 °C.

2.2. Amplification of bacterial 16S rDNA and glnA gene fragments

Frankia genomic DNA was purified from plant root nodules or cultured strains as previously described (Clawson et al., 1998). Portions of Frankia 16S rDNA and glutamine synthetase (glnA) genes were amplified from the same DNA preparations. Polymerase chain reaction (PCR) primers used for the 16S rRNA gene were fD1 (Weisburg et al., 1991) and rDB1 (Benson et al., 1996). The glnA PCR primers were designated DB41, DB42, and DB44.

Table 1 lists the phylogenetic groups, families and genera of actinorhizal plants.

<table>
<thead>
<tr>
<th>Phylogenetic groups of “core rosids” (Eurosids I sensu APG)*</th>
<th>Familyb</th>
<th># Genera/# nod.</th>
<th>Genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Higher” Hamamelididae (Fagales sensu APG)</td>
<td>Betulaceae</td>
<td>6/1</td>
<td>Ahus</td>
</tr>
<tr>
<td></td>
<td>Casuarinaceae</td>
<td>4/4</td>
<td>Allocasuarina, Casuarina, Ceuthostoma, Gymnostoma</td>
</tr>
<tr>
<td></td>
<td>Myricaceae</td>
<td>3/2</td>
<td>Comptonia, Myrica</td>
</tr>
<tr>
<td>Urticales (included in Rosales sensu APG)</td>
<td>Elaeagnaceae</td>
<td>3/3</td>
<td>Elaeagnus, Hippophae, Shepherdia</td>
</tr>
<tr>
<td></td>
<td>Rhamnaceae</td>
<td>55/7</td>
<td>Ceanothus, Colletia, Discaria, Kentrothamnus, Retanilla, Talguenea, Trevoa</td>
</tr>
<tr>
<td>Rosaceae (included in Rosales sensu APG)</td>
<td>Rosaceae</td>
<td>100/4</td>
<td>Chamaebetia, Cercocarpus, Dryas, Purshia (includes Covania)</td>
</tr>
<tr>
<td>Cucurbitales (Cucurbitales sensu APG)</td>
<td>Coriariaceae</td>
<td>1/1</td>
<td>Coriaria</td>
</tr>
<tr>
<td></td>
<td>Datiscaeace</td>
<td>3/1</td>
<td>Datisca</td>
</tr>
</tbody>
</table>


**Classification according to Cronquist (1981).**

*Covania has been included with Purshia (Henrickson, 1986).**
parsimony phylogenetic trees with bootstrap values were generated from the alignments in PHYLIP (Felsenstein, 1993). Five hundred pseudo-alignments were generated with Seqboot. Distance matrices were constructed in DNAdist with the Kimura two-parameter model (transition/transversion ratio of two). Neighbor-Joining trees were constructed with the program Neighbor (Saitou and Nei, 1987). A consensus tree was generated with Consense and viewed as a cladogram in TREEVIEW (Page, 1996). Parsimony trees were generated in PHYLIP with Dnapsars from the pseudo-alignments described above. Maximum likelihood trees were produced from glnA alignments with either the second codon position or full-length sequence, 16S rDNA alignments, and concatenated alignments of glnA and 16S rDNA sequences in TREE-PUZZLE (Strimmer and Haeseler, 1996) with the HKY model of substitution (Hasegawa et al., 1985). The alignments were tested for clock-like rates of mutation with the molecular clock option in TREE-PUZZLE.

3. Results

3.1. Bacterial DNA sequence

Reported here are nine partial Frankia 16S rDNA sequences including the first from the rosaceous genera Cercocarpus and Chamaebatia, 24 partial glnA sequences, and the partial glnA sequence of *A. cellulolyticus*. We used *A. cellulolyticus* as an outgroup in this study because previous phylogenetic work has shown it to be the closest known relative to *Frankia* (Marechal et al., 2000; Normand et al., 1996). The partial sequences of Frankia 16S rRNA genes used in this study correspond to positions 28–419 on *Escherichia coli* 16S rDNA; alignments contain 44 variable sites in 375 bp (11.7% variability in the region), or about 45% of the total number of variable sites of the full-length Frankia gene (Clawson et al., 1998). The portion of the glnA gene used spans 477 nucleotides from a total of about 1500 with 162 informative sites among the Frankia sequences (33.9%); this region contains a conserved 13 amino acid region involved in active site formation (Hosted et al., 1993). An alignment of GS amino acid residues revealed 30 variable positions from a total of 159 (data not shown).

3.2. Frankia phylogeny

The phylogeny of Frankia based on a data set of concatenated 16S rDNA and glnA sequences is shown in Fig. 1. The topology of the tree was strongly supported by Neighbor-Joining, Maximum Likelihood and Maximum Parsimony and is in agreement with previous studies by placing the strains in three major clades. In all trees, Clade II clusters with Clade III with high bootstrap support (92, 100, 81 for Neighbor-Joining, Maximum Likelihood, and Parsimony, respectively).

Sequences from unisolated Frankia strains of the Rosaceae genera Cercocarpus, Chamaebatia and Parshia...
Cluster together in Clade I along with unisolated strains from Datisca (Datispaceae), Coriaria (Coriariaceae) and Ceanothus (Rhamnaceae). Clade II contains only Frankia symbionts and strains from the “higher” hamamelid lineage. Clade III contains Frankia strains from the Urticales lineage plus strains R2, isolated from Ceanothus americanus and PtI6, isolated from Parshia tridentata. The latter two strains represent frankiae that fail to reinfect (Nod−) the host of origin (Benson and Silvester, 1993). One of the strains (R2) has previously been shown to effectively infect Elaeagnus sp. in greenhouse trials (Baker, 1987). The ability of the other strain used in this study (PtI6) to infect Elaeagnus has not been reported but we may predict, on the basis of this study, that it has a high probability of doing so.

The presence of clock-like mutation rates in DNA or protein sequence alignments is helpful for correlating time within evolutionary trees. The maximum likelihood ratio test (lrt) can be used to test for molecular clocks in sequence alignments (Huelsenbeck and Rannala, 1997). The null hypothesis of the lrt is clock-like evolution (Huelsenbeck and Rannala, 1997). Thus, p scores less than 0.05 reject a molecular clock. In agreement with a previous study (Jeong and Myrold, 1999), the lrt rejected clock-like evolution in our alignment of Frankia 16S rDNA sequences (p = 0.0094). To test for a molecular clock that would be useful in calibrating our glnA phylogenetic trees, we first combined our glnA alignment with the following sequences: Rhizobium leguminosarum bv. viciae (AF169566), Sinorhizobium
The molecular clock hypothesis could not be rejected when the conserved second codon position of the alignment was analyzed ($p = 0.0536$). The resulting tree from that analysis is shown in Fig. 2.

Since all taxa in Fig. 3 adhere to clock-like behavior, nodes of the entire tree can be calibrated to the divergence of E. coli and S. typhimurium. The molecular clock hypothesis could not be rejected when the conserved second codon position of the alignment was analyzed ($p = 0.0536$). The resulting tree from that analysis is shown in Fig. 2.

The combined data sets for 16S rRNA and glnA genes provide a well-supported phylogeny for effective Frankia strains. The basal position of Clade I in the tree presented in Fig. 1 is found in each method of analysis. The tree is the most comprehensive to date and includes data from representative strains of all known actinorhizal plant genera except for Allocasuarina, Cesthostoma, Dryas, Gymnostoma, Kentrothamnus, and Shepherdia. However, the phylogenetic positions of Frankia sp. strains inhabiting nodules of the majority of plants from these genera are known (Bosco et al., 1994; Navarro et al., 1997; Nazaret et al., 1991; Simonet et al., 1999). The information provided allows for a more complete picture of plant–microbe relationships than previously available and provides the basis for hypotheses concerning the evolution of the symbiosis.

Fig. 3 summarizes the relationships between Frankia clades and actinorhizal plant genera from the “core Rosids.” Phylogenetic groups of the plants are shown on the left, Frankia clades are shown on the right. Solid lines indicate the presence of typical Frankia strains in each plant genus. Broken lines indicate that a Frankia clade is only occasionally detected in the plant genus. Thick lines denote root infection via root hairs; thin lines denote infection via intercellular penetration of the root surface.

4. Discussion

The combined data sets for 16S rRNA and glnA genes provide a well-supported phylogeny for effective Frankia strains. The basal position of Clade I in the tree presented in Fig. 1 is found in each method of analysis. The tree is the most comprehensive to date and includes data from representative strains of all known actinorhizal plant genera except for Allocasuarina, Cesthostoma, Dryas, Gymnostoma, Kentrothamnus, and Shepherdia. However, the phylogenetic positions of Frankia sp. strains inhabiting nodules of the majority of plants from these genera are known (Bosco et al., 1994; Navarro et al., 1997; Nazaret et al., 1991; Simonet et al., 1999). The information provided allows for a more complete picture of plant–microbe relationships than previously available and provides the basis for hypotheses concerning the evolution of the symbiosis.

Fig. 3 summarizes the relationships between Frankia clades and actinorhizal plant genera from the “core Rosids.” Phylogenetic groups of the plants are shown on the left, Frankia clades are shown on the right. Solid lines indicate the presence of typical Frankia strains in each plant genus. Broken lines indicate that a Frankia clade is only occasionally detected in the plant genus. Thick lines denote root infection via root hairs; thin lines denote infection via intercellular penetration of the root surface.
Additional specialization, or coevolution, has been reported at lower taxonomic levels within the hamamelids and may be linked to geographical separation (Caizzi et al., 1990; Simonet et al., 1999). Clade III Frankia strains are effective symbionts of two of the four lineages including the actinorhizal Urticales and the “higher” Hamamelidae (Carú, 1993; Clawson and Benson, 1999; Clawson et al., 1998; Navarro et al., 1997). They are the dominant strains in the Elaeagnaceae and, except for Ceanothus, in the nodulated Rhamnaceae. However, they have been isolated from Purshia sp. (Rosaceae, strain Pt16 in Fig. 1) and Ceanothus sp. (Rhamnaceae, strain R2 in Fig. 1) (Lechevalier and Ruan, 1984), and their 16S rDNA has been detected in nodules of Ceanothus sp. (Murry et al., 1997). Some of these strains (such as R2) nodulate Elaeagnus sp. in greenhouse trials (Baker, 1987). Among the “higher” hamamelids, Clade III strains are common in Myrica and Gymnostoma nodules and can induce ineffective (non-nitrogen-fixing) nodules on alders (Bosc et al., 1992; Clawson and Benson, 1999; Margheri et al., 1985; Navarro et al., 1997). Some have also been isolated from Casuarina sp. (Diem et al., 1982; Gauthier et al., 1999; Navarro et al., 1997). Strains from Clade III therefore inhabit nodules of actinorhizal plants from three major clades, although their roles in the Rosaceae, in Ceanothus of the Rhamnaceae, and in some genera of the “higher” hamamelids are unclear.

The divergence of Frankia clades estimated in this study precedes prior estimates based on either nitrogenase (nifH) or 16S rRNA sequence analyses. The divergence of Clade I from the ancestor of II and III has been estimated at 62–130 Myr bp (Jeong et al., 1999) and 100–200 Myr bp (Normand et al., 1996), and the divergence of Clade II from Clade III at 46–91 Myr (Jeong et al., 1999), 25–50 Myr (Normand et al., 1996), 30 Myr (Normand and Bousquet, 1989), and 131 ± 56 Myr (Cournoyer et al., 1993). Our analysis with glnA sequences indicates that all three Frankia clades emerged well before pollen similar to present day members of the higher hamamelids appears in the fossil record around 95 Myr bp (Sims et al., 1999) and even before angiosperm fossils appear in the Cretaceous period about 125 Myr bp (Crane et al., 1995).

The dating of Frankia clades before the appearance of angiosperms in the fossil record obscures the history of coevolution between actinorhizal plants and their symbionts. Current associations of actinorhizal plants and Frankia strains indicate non-congruent phylogenies (Fig. 3) that may be the result of either multiple gains or losses of the symbiosis. Some families of actinorhizal plants have a relatively shallow fossil history. Some of the lineages have no fossil predecessors before a few tens of millions years ago despite molecular evidence that suggests that all lineages were established at about the same time in the late Cretaceous (Magallon et al., 1999). Clearly, the lack of identifiable fossils from an actinorhizal lineage does not mean that the symbiosis did not exist prior to the deposition of fossils. Likewise, the absence of any recognizable actinorhizal lineage in the mid-Cretaceous does not mean that the ancestors of Frankia were not participating in plant symbioses before the lineages became established. In effect, it is difficult if not impossible to draw strong conclusions about the early evolution of the actinorhizal symbiosis.

A comparison of actinorhizal plant and Frankia clades indicates that intercellular penetration is the likely ancestral mode of plant infection since Clades I and III Frankia strains share that ability and infect a broader range of plants than Clade II strains (Fig. 3). When ancestors of Clade II specialized in infecting ancestral hamamelids via root hair infection they apparently lost the ability for intercellular penetration and subsequently coevolved with the “higher” hamamelids. This scenario differs from one proposed previously where the divergence of Clades II and III was correlated with the evolution of the intercellular penetration mode of infection when the Elaeagnaceae lineage appeared in the fossil record (Cournoyer and Lavire, 1999; Cournoyer et al., 1993; Normand et al., 1996).

In sum, the present work helps clarify the phylogeny of Frankia by utilizing a large collection of strains and an informative region of genomic DNA, the glnA gene. The evolution of three distinct Frankia phylotypes (Clades I–III) is predicted to predate the fossil history of extant plants. Thus, the early evolution of the actinorhizal symbiosis is obscured by an incomplete fossil record. While 16S rDNA sequences are still useful for placing strains in broad phylogenetic groups, the inclusion of glnA sequences allows the grouping to be done with a higher degree of confidence.

Acknowledgments

This project was supported by the United States Department of Agriculture, NRICGP. We thank Dave Nelson of the Provo Shrub Sciences Laboratory of the United States Forest Service for supplying nodules from members of the Rosaceae.

References


