Analysis of *Frankia* evolutionary radiation using *glnII* sequences

Benoit Cournoyer*, Céline Lavire

Laboratoire d'Ecologie Microbienne des Sols, UMR CNRS 5557, Université Claude-Bernard Lyon 1, 69622 Villeurbanne Cedex, France

Received 1 February 1999; received in revised form 5 May 1999; accepted 27 May 1999

**Abstract**

Using a *glnII* (encoding glutamine synthetase II) PCR selective screening, a *Frankia* ACN14a gene library clone was isolated. A derived *glnII*-hybridising 2.7-kb *Hind*III subclone was characterised. Identities of 95% and 93% were observed, respectively, with the corresponding *Frankia* CprI1 *glnI* and *glnII* regions. A variable segment of the *glnII* region was selected, PCR amplified from various *Frankia* genomes, sequenced, and used to investigate phylogenetic relationships within the genus. *glnII* phylogenetic inferences are well-resolved and allowed us to deduce evolutionary trends among *Frankia*. *Frankia* radiation seems to begin with a diversification according to the ability or not to infect actinorhizal plants. The infective strains are divided into two clusters matching plant-colonising specificities. © 1999 Published by Elsevier Science B.V. All rights reserved.

**Keywords:** Glutamine synthetase; Symbiosis; Nodule; Actinomycete; Evolution

---

**1. Introduction**

The genus *Frankia* includes both (a) actinomycetic, nitrogen-fixing, nodule-forming endophytes of actinorhizal plants such as *Alnus*, *Myrica*, *Casuarina*, *Elaeagnus*, and *Hippophae* and (b) free-living actinomyces without nodule-forming and nitrogen-fixing capacity but producing as in group (a) sporangia containing nonmotile spores in submerged liquid culture [1]. *Frankia*-infective isolates can be divided into two major groups according to host plant-specific responses to invasion. One group is adapted for root hair infection (RHI) involving micro-symbiont intracellular migration toward a nodule primordium [2], and contains strains infective on *Alnus*, *Myrica* and *Casuarina*. The other group is adapted for direct root penetration (IP) involving intercellular migration of the micro-symbiont toward a nodule primordium [3], and contains strains infective on *Elaeagnus* and *Hippophae*.

Recent data about the molecular phylogeny (using *nif* and *rrn* loci) of the genus *Frankia* evidenced a close relationship between the speciations and symbiotic properties of these micro-organisms [4]. However, many aspects of the *Frankia* evolutionary tree remained unclear. Here, we investigated the use of glutamine synthetase encoding genes to improve our understanding of the phylogenetic relationships among *Frankia*. A *Frankia* ACN14a gene library was screened to isolate a *glnII*/*glnII* library clone. *glnII*/*glnII* DNA sequences were obtained from this
clone and compared with those previously reported for the closely related Frankia CpI1 strain. glnII showed greater variability and was chosen to investigate Frankia phylogenetic trends.

2. Materials and methods

2.1. Bacterial strains and DNA extraction

The following Frankia representative strains were used in this study: ArI3 [5], ACN1AG [6], and ARgN22d [7] of Frankia alni (Alnus infectivity group, RHI); ACN14a [8], AVN17o [9] and ARgP5AG [7] (Alnus infectivity group, RHI); M2 [10] (Casuarina infectivity group, RHI); HRN18a [11] (Elaeagnus infectivity group, IP) and EUN1f [6] (Elaeagnus infectivity group, IP but also infective on Alnus, RHI); PtI1 [12] (non-infective and non-nitrogen-fixing Frankia group). Growth conditions for Frankia isolates and DNA extraction procedures were according to Simonet et al. [13].

2.2. Frankia gene library, glnII sub-cloning and gene sequence analyses

Frankia ACN14a gene library was built according to standard procedures [14]. Plasmid pRK404 [15] was chosen as cloning vector, cut by BamHI and dephosphorylated. Frankia DNA was cut partially using Sau3AI (Boehringer Mannheim, France). DNA fragments were size-fractionated by sucrose gradient. Electro-transformation and selection of transformants were performed according to the manufacturer of the Esherichia coli DH10B strain (Gibco BRL/Life Technologies, France). DNA manipulations were performed following standard procedures [14]. DNA manipulation analyses were performed using the GCG package of the University of Wisconsin [16] and the Bisance package (Université Paris V, France). Blast searches [17] were performed to detect DNA and protein similarities.

2.3. PCR amplification and sequencing of glnII genes

DNA amplification by PCR (100 µl) was performed according to the Taq polymerase manufacture (Gibco BRL/Life Technologies, France). PCR cycles were as follows: (1) 95°C for 5 min, (2) 95°C for 1 min, 57°C for 1 min, and 72°C for 1 min (35 cycles), and (3) 5 min at 72°C. Primers Fggs19 (5'-TAC ATCTGGATCCAGCACG-3') with either FGgs338 (5'-GAAGGTTGACCCCTTGCTCGA-3') or FGgs417 (5'-GCGGACGCAGCTAGTA-3') were used to amplify regions of glnII. PCR amplification between Fggs19 and FGgs417 and between Fggs19 and FGgs338 are expected to match nucleotide positions 19–338 and 19–417 of the Frankia CpI1 coding region [18]. Direct sequencing of PCR-amplified DNA was done according to Winship [19].

2.4. Phylogenetic analysis

glnII DNA sequences were aligned using the multiple alignment clustalv software [20]. Alignments were corrected to match gene codons. Sites involving indels (insertions/deletions) were excluded from all analyses. Evolutionary distances between sequence pairs corrected for multiple substitutions were computed using Kimura’s two-parameter model [21]. The phylogenetic tree was inferred using the neighbour-joining (NJ) method [22]. Bootstraps were performed [23].

3. Results and discussion

3.1. Characterisation of a glnII Frankia library clone

A Frankia ACN14a gene library was built. About 2500 transformants were picked individually and transferred to 96-well microtitre plates. These clones were pooled and analysed using a PCR glnII-selective screening (using primers Fggs19 and FGgs417). Five pools of 5×96 clones were prepared and four of these gave a single positive PCR product of the expected size (about 395 bp) suggesting that an equivalent of four Frankia genomes were cloned. One of these positive pools was divided into five pools of 96 clones. A single PCR product was detected for one of these. This positive pool was divided into 12 pools of eight clones. These were screened by PCR, allowing the identification of the positive pool of eight clones. Individual clones of this pool were then
tested and a single clone carrying an insert of about 17 kb was identified and named pFQ310 (Fig. 1).

DNA blot analysis was carried out on pFQ310 using a *Frankia* ACN14a *glnII* DNA probe (obtained by PCR using primers FGgs19 and FGgs417). A hybridising 2.7-kb *HindIII* fragment was observed and subcloned into pBluescript (KS−). DNA sequencing was performed from both ends of this clone, which was named pFQ313, using the universal T3 and T7 (EcoRI side of the insert) primers. 815 bp of the T7 and 573 bp of the T3 ends were sequenced. Blast searches confirmed the presence of a *glnII* sequence at the T7 end and detected a *glnI* sequence at the T3 end. Comparisons were performed with the *Frankia* CpI1 reported *glnI* [24] and *glnII* [18] DNA sequences: 93% identity was observed between the *glnII* regions and 95% between the *glnI* regions. Analyses of the ends of the sequenced regions showed the presence of a segment of the pRK404 polylinker region at the T7/*glnII* end. This observation showed that pFQ310 like pFQ313 encoded only a partial *glnII* sequence.

Putative promoters, ribosome binding sites and start codons were detected upstream of *glnI* and *glnII* sequenced regions. GCG Frames and Testcode analyses were performed, respectively, to identify putative ORF (open reading frames) and to estimate the coding bias of the reported DNA sequences. The Testcode analysis of the *glnI* sequence (GenBank accession number AF121818) suggested the presence of a highly probable coding sequence from position 200 (corresponding to a GTG codon). From position 120 to 200, the signals were variable being sometimes highly probable or less probable coding sequences. The putative start codon is positioned at nucleotide 178 [24]. Upstream of nucleotide position 120, the DNA sequence does not seem to be a coding sequence. Frames analysis confirmed the presence of an ORF from position 178 of the *glnI* sequence. A single *Frankia*-E. coli type promoter was detected upstream of the *glnI* region. Its position (−10 around position 146–158; −35 around position 115–130) fits with the putative ribosome binding site (RBS) observed and the ATG codon at position 178. The G+C% of the sequenced region is 59%.

Testcode analysis of the *glnII* sequence (GenBank accession number AF121819) suggested the presence of a highly probable coding sequence from position 185. At position 185, no start codon could be observed but a putative RBS region was found very close to it (nucleotides 188–191). The putative GTG start codon [18] maps at position 198 and fits well with this highly probable coding sequence. Upstream of nucleotide 185, the likelihood of the sequence to be coding decreases rapidly from position 185 to 170. Position 170 corresponds to a segment of the putative Ntr promoter [18]. Upstream of nucleotide 170, the sequence is unlikely to be coding. Frames analysis confirmed the position of an ORF which could begin either from the classical ATG codon at position 264 or from the GTG codon at position 198. One main difference, which is observed between strains ACN14a and CpI1 *glnII* sequences, is the presence of indels that led to five codons (four of which are at the beginning of the ORF) that are not found in the CpI1 sequence. The G+C% of the sequenced *glnII* region is 67%.

3.2. *Frankia* genomic species *glnII* PCR amplification and DNA sequences

The *Frankia* ACN14a sequenced *glnII* region shows greater variability than the *glnI* region according to their similarities with *Frankia* CpI1 sequences. The *glnII* region was thus chosen to investigate phylogenetic relationships among *Frankia* strains. PCR
(from primers FGgs19 and FGgs417) was used to amplify a glnII region from various Frankia genomes. The glnII sequences obtained from the amplified fragments (around 400 bp) are about 200 bp (see Fig. 2 for the strains analysed and the respective accession numbers). The multiple alignment derived from these sequences contains 75 variable sites including 44 cladistically informative ones. Previous nif and rrn multiple alignments used to investigate Frankia phylogeny showed, respectively, 45 (28 informative) and 14 (13 informative) variable sites [4]. The glnII alignment is thus more informative than these previous alignments. The glnII segment of all Frankia strains sequenced and of the Streptomyces viridochromogenes sequence (GenBank accession number X52842) are extremely G+C-rich: their overall G+C content ranges from 67 to 71%, while the G+C content of their third codon positions varies between 92 and 97%.

3.3. Molecular phylogeny of the glnII gene

The molecular phylogeny of Frankia glnII DNA is well resolved (Fig. 2). This is the first phylogeny which positions at significant levels the main lineages observed among Frankia. The non-infective and non-nitrogen-fixing Frankia representative strain (PtI1) glnII sequence is positioned outside the evolutionary radiation of the symbiotic Frankia glnII sequences for 95% of the bootstrap replicates. The symbiotic Frankia glnII sequences are divided at a statistically significant extent into two clusters: (1) a cluster grouping glnII sequences of the representative IP-infective strains i.e. EUN1f and HRN18a (for 98% of the bootstrap replicates), and (2) a cluster grouping glnII sequences from the strictly RHI-infective strains i.e. M2, ARgN22d, ACN14AG, ArI3, CpI1, ACN14a, AVN17o and ARgP54G strains (for 99% of the bootstrap replicates). Within cluster 2, the ACN14a glnII phylogenetic lineage is shown to diverge significantly (for 100% of the bootstrap replicates) from the F. alni strains (ARgN22d, ACN14AG, CpI1 and ArI3) glnII lineage. This divergence may be related to the loss of endogenous plasmids by ACN14a. The strain ACN14a does not possess endogenous plasmids but the F. alni strains studied here do [13].

Previous nif and rrn phylogenies of the genus have been used to resolve parts of the Frankia evolutionary tree but many aspects remained unclear. Using 16S ribosomal DNA sequences [4, 25], the IP-infective strains were shown to group together but it re-
mained ambiguous whether the IP strains came from the strictly RHI-infective ones or from a point of the radiation reflecting a specialisation that led to a diversification of the infective strains into an IP- and a strictly RHI-infective lineage. The *Casuarina*-infective strains of the RHI group could not be positioned accurately using the 16S rDNA data set. Using *nif* sequences, few clusters could be resolved significantly but a combined data set of *nif* and 16S rDNA showed *Casuarina*-infective strains to be closely related to *F. ahi* [4]. From this work, it is clear that the infective strains diverged at a point of the radiation to lead to the emergence of two lineages: (1) the IP-infective strains, and (2) the strictly RHI-infective strains. It also appears that the non-infective *Frankia* strains (represented by PtI1) diverged from the infective ones prior to their division according to the infection process and host specificities. This raises interesting questions about a hidden (and poorly investigated) diversity among non-infective *Frankia* strains. From the complexity of the infective cluster, it is likely that the non-infective strains will show as much diversity and complexity.

A speculative scenario about the evolution of *Frankia* symbiotic properties can be drawn from the *glnII* phylogeny. The radiation seems to begin by a separation according to the ability or not to infect actinorhizal plants. Then, a first adaptation of *Frankia* symbiotic properties toward infection of *Myrica* plants (or ancestors) by the RHI process seems likely. A species of *Myrica* was shown to be a promiscuous host plant that can be infected by *Frankia* from all groups of plant specificity [12]. These *Myrica* infective strains would then have adapted their properties for the infection of *Alnus, Elaeagnus* or both (or their ancestors). Infection of *Elaeagnus*-related plants would have been achieved through the emergence of the IP infection process. This adaptation would be reflected in the *Frankia* phylogeny by a division of the strains according to the infection process, i.e. strictly RHI-infective or IP-infective. Strains infective on *Casuarina* (RHI) would have emerged from strains infective on *Alnus* (strictly RHI). This adaptation suggests a change in the infection process which became incompatible with the infection of *Alnus* (*Casuarina*-infective strains do not infect *Alnus* spp.).

References


